Technical Report 1334

Best Practices and Provisional Guidelines for Integrating Mobile, Virtual, and Videogame-Based Training and Assessments

Robert C. Brusso
Robert A. Wisher
Arthur Paddock
Joshua Hatfield
ICF International

January 2014

United States Army Research Institute for the Behavioral and Social Sciences

U.S. Army Research Institute
for the Behavioral and Social Sciences

Department of the Army
Deputy Chief of Staff, G1

Authorized and approved for distribution:

MICHELLE SAMS, Ph.D.
Director

Research accomplished under contract
for the Department of the Army by

ICF International

Technical review by

Michael Singer, U.S. Army Research Institute
Marisa Miller, U.S. Army Research Institute

NOTICES

DISTRIBUTION: Primary distribution of this Technical Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: DAPE-ARI-ZXM, 6000 6th Street (Bldg. 1464 / Mail Stop: 5610), Ft. Belvoir, Virginia 22060.

FINAL DISPOSITION: Destroy this Technical Report when it is no longer needed. Do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Technical Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy)	2. REPORT TYPE	3. DATES COVERED (from. . . to)
January 2014	Final	December 2011 – December 2012

4. TITLE AND SUBTITLE	5a. CONTRACT OR GRANT NUMBER
Best Practices and Provisional Guidelines for Integrating Mobile, Virtual, and Videogame-Based Training and Assessments	W5J9CQ-11-D-0002
	5b. PROGRAM ELEMENT NUMBER 622785
6. AUTHOR(S) Robert C. Brusso, Robert A. Wisher, Arthur Paddock, and Joshua Hatfield	5c. PROJECT NUMBER A 790
	5d. TASK NUMBER
	5e. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) ICF International 9300 Lee Highway Fairfax, VA 22030	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 6000 6th Street (Bldg. 1464 / Mail Stop: 5610) Fort Belvoir, VA 22060	10. MONITOR ACRONYM ARI
	11. MONITOR REPORT NUMBER Technical Report 1334

12. DISTRIBUTION/AVAILABILITY STATEMENT:

Distribution Statement A: Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Subject Matter POC and Subject Matter Expert: Randall D. Spain

14. ABSTRACT *(Maximum 200 words)*:

The Army needs guidance on how to use and integrate assessments and various learning technologies to achieve and support the principles set forth in the Army Learning Model. In an initial step to provide best practices, a literature review and interviews were conducted with subject matter experts to find exemplars within the U.S. Army, other U.S. Armed Services, or other nations' Armed Forces, and the private sector in which training and assessment has been carried out operationally using the following platforms: mobile devices, virtual worlds, and videogame-based scenarios. From the literature review, three exemplars and 23 exemplary elements were found. The exemplars and exemplary elements, along with insights from the interviews, were analyzed to identify several best practices and guidelines for training and assessment development. Given the infancy of the research in this area, however, the practices and guidelines identified can only be considered provisional until more evidence can be gathered.

15. SUBJECT TERMS

Army Learning Model, Assessments, Guidelines, Mobile training, Virtual training, Videogame-based training

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unclassified Unlimited	73	Dorothy Young 703-545-4225

Technical Report 1334

Best Practices and Provisional Guidelines for Integrating Mobile, Virtual, and Videogame-Based Training and Assessments

Robert C. Brusso, Robert A. Wisher, Authur Paddock, and Joshua Hatfield
ICF International

Orlando Research Unit
Joan H. Johnston, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
6000 6th Street, Bldg. 1464
Fort Belvoir, VA 22060

January 2014

BEST PRACTICES AND PROVISIONAL GUIDELINES FOR INTEGRATING MOBILE, VIRTUAL, AND VIDEOGAME-BASED TRAINING AND ASSESSMENTS

EXECUTIVE SUMMARY

Research Requirement:

The Army Learning Model (ALM) contains a guiding set of principles for Army learning practices that focus on anytime, anywhere training with a learner-centered approach to enable Soldiers to learn faster and adjust more quickly to complex and uncertain environments (TRADOC, 2011). To achieve these goals, the ALM proposes the use of learning technology, in addition to face-to-face instruction, and assessment. However, the Army needs guidance on how to use and integrate assessments and various learning technologies in a way that will meet the goals of the ALM.

In response to these needs, the U.S. Army Research Institute for the Behavioral and Social Science (ARI) developed prototype training applications and assessments to serve as a test-bed for conducting research on assessment strategies in maturing learning technologies, including mobile devices, virtual classrooms, and collaborative game-based technologies. In an initial step to guide the development of the prototype training, we sought to find exemplars within the U.S. Army, other U.S. Armed Services, or other nations' Armed Forces, and the private sector in which training and assessment has been carried out operationally using the following platforms: mobile devices, virtual worlds, and videogame-based scenarios. From these exemplars, best practices for integrating mobile, virtual and videogame-based training, and for using and administering assessments through these different platforms, along with benefits and challenges involved with these technologies, were identified. From the best practices, a set of guidelines for future training and assessment development were created.

Procedure:

A thorough literature review was conducted and training and assessment experts were interviewed to identify exemplars and best practices for integrating assessments within the platforms of interest (i.e., mobile devices, virtual worlds, and videogame-based training). Approximately 1,200 sources were found through a search of online databases (e.g., Education Resources Information Center, PsycINFO, and Defense Technical Information Center) and electronically available conference proceedings using a variety of search terms. The abstracts of the 1,200 sources to determine if the source provided an evaluation of a training that included one or more of the platforms of interest, implementation or development of an assessment or training evaluation process, or discussed lessons learned regarding the platforms of interest or assessments. Seventy-seven sources included one or more of these elements and were retained for full coding. Following the coding, the coded information was reviewed for each of the 77 sources to determine if the source contained an exemplar. Sources that did not meet the exemplar criteria were reviewed for "exemplary elements"- insightful approaches or recommendations regarding the use or integration of the platforms of interest or the use of assessments that were unique or innovative and represented an effective practice.

In addition to the literature search, 22 subject matter experts (SMEs) were interviewed that met one of the following criteria: 1) have specific knowledge of assessment methods within learning technology platforms, 2) be a leading researcher or academician with a known research and publication record within the domains of training and assessment (i.e., publication productivity in refereed journals), 3) have extensive knowledge of training implementation across relevant learning technology platforms, or 4) have knowledge of specific potential exemplars. Notes from each interview were reviewed and analyzed for key comments that provided insight about the implementation or development of an assessment or evaluation process, or the integration or use of the platforms of interest.

Following the literature review and interviews, exemplars, exemplary elements, and interview insights were analyzed to identify discernible patterns or consistent themes. A set of provisional guidelines and proposed practices were developed in accord with the themes in each category.

Findings:

Of the 77 sources, 3 contained exemplars and 23 contained exemplary elements. None of the three exemplars utilized all three of the training platforms that are targeted for the prototype training, but each did employ multiple training platforms, with at least one or more of those platforms being a mobile, virtual world or videogame platform. All three exemplars included a pre-test and one or more post-tests, and all three included measures of learning outcomes, which were either multiple-choice or situational judgment tests. The 23 exemplary elements that were identified described unique or innovative approaches that represented an effective practice. The remaining sources did not meet the criteria for exemplars and did not provide an approach to integrating the platforms of interest or to the use of assessments that could be considered unique, innovative, and/or an effective practice.

Fourteen themes were identified within the categories of: 1) best practices for integrating mobile, virtual and videogame-based platforms, 2) role of assessments and how they can be implemented within these platforms, or 3) benefits or challenges of the platforms. In accord with the themes, the following provisional guidelines were developed:

- Guideline 1: To ensure the effectiveness and efficiency of training, apply principles of the learning sciences in order to plan the integration of learning experiences across training platforms.
- Guideline 2: Develop an assessment strategy and incorporate that strategy into the training framework to ensure assessment(s) provide learner performance data that support the overall goals of the training.
- Guideline 3: Use the same platform for assessment that was used for training, if the platform can adequately capture the necessary assessment data, to maximize training efficiency.
- Guideline 4: Use assessments to adapt training to learners' proficiency levels and/or style preferences to support a learner-centric environment.
- Guideline 5: Employ frequent testing as a means to deliver content, reinforce what was learned, and support adaptive instruction.

- Guideline 6: Implement a holistic method for aggregating and analyzing all sources of social exchange data to assess critical learning objectives within peer- and collaborative-learning scenarios.
- Guideline 7: Capitalize on the capabilities of the training technology to assess competencies in alternative ways that are not feasible through traditional platforms.

Several proposed practices were also developed. Given the infancy of the research in this area, the practices and guidelines can only be considered proposed or provisional until more evidence can be gathered.

Utilization and Dissemination of Findings:

The provisional guidelines, combined with well-established guidelines for distance learning and learner-centered approaches, provide a foundation from which to build the future ALM. The findings from this qualitative research provide initial answers to the research questions ARI seeks to address in the development of the prototype training. The findings also provide an empirical baseline for identifying areas for future research.

BEST PRACTICES AND PROVISIONAL GUIDELINES FOR INTEGRATING MOBILE, VIRTUAL, AND VIDEOGAME-BASED TRAINING AND ASSESSMENTS

CONTENTS

APPENDICES

Page

TABLES

BEST PRACTICES AND PROVISIONAL GUIDELINES FOR INTEGRATING MOBILE, VIRTUAL, AND VIDEOGAME-BASED TRAINING AND ASSESSMENTS

Introduction

The new Army Learning Model (ALM) as described in the U.S. Army Learning Concept for 2015 sets forth an ambitious agenda for innovation in Army training (TRADOC, 2011). It contains a guiding set of principles for Army learning practices to enable Soldiers to learn faster and adjust more quickly to complex and uncertain environments than potential adversaries, and create a fighting force that exhibits a high degree of operational adaptability in an era of persistent conflict (TRADOC, 2011). These principles focus on the importance of anytime, anywhere training that can actively engage learners, from recruits to retirees, with a learner-centered approach.

Anytime, Anywhere Training

Rather than limiting training to specific timeframes and locations (e.g., a 'brick and mortar' training environment), the ALM emphasizes that the training system must be accessible at the 'point of need.' At the 'point of need' refers to both the ability to be immediately accessible from any location or at any time of the day, and the ability to address the needs that the learner has at that moment. As such, a high level of importance is placed on having training that is distributed and flexible so that it "…extends knowledge to Soldiers at the operational edge, is capable of updating learning content rapidly, and is responsive to Operational Army needs" (TRADOC, 2011, p. 16).

Learner-Centered Approach

To support the individual learning experience and actively engage learners, many educational technologists advocate the need to shift from instructor-centered to learner-centered teaching approaches. Current Army training is principally instructor-led and not synchronized to meet individual learner needs, which will vary depending on operational environment, performance, and goals (e.g., career goals, assignment goals). Learner-centered pedagogy asks what students need to learn, what their learning preferences are, and what is meaningful to them. It ensures that training is tailored to the individual learner's level of experience and competence. The term andragogy is sometimes used to refer to learning strategies applied to adults (Knowles, 1980), and is relevant here because some of the commonly accepted principles of andragogy align with a learner-centered approach, to include facilitating self-directed and autonomous learners, and the role of trainers and teachers as supporting and facilitating this dynamic (Knowles, 1980).

Use of Learning Technology

To achieve these outcomes, the ALM proposes the use of advances in learning technology, such as those related to virtual training environments, in addition to face-to-face instruction. This blended learning approach is expected to capitalize on efficiencies of technology-based instruction, realize the advantages of ubiquity that these methods allow (i.e.,

offering constant availability to Soldiers through mobile, and computer-based platforms), and allow training to be individualized and adapted to the learner (e.g., accounting for varying skill level and pace of instruction). The blended approach also leverages the competence of digital age learners in using technology, as it is proposed that younger Soldiers have been raised in an environment where they are familiar with, and have facility using emerging technologies (e.g., video-games, Web 2.0 capabilities). The argument is that Army training methods should use advanced technology that is commonly used by the majority of Soldiers, and should support continued development of competence in the use of technology (as use of technology is now a key component of most Army jobs).

The use of learning technology can allow a training delivery system to be flexible enough to quickly adapt to changes in the learner so that instruction can be tailored in an individualized manner (e.g., by adapting content to the learner's aptitude level). In this case, training is more relevant to the learner, supports motivation to learn and promotes self-directed learning. In addition, the inclusion of rich, multi-media training tools is anticipated to increase motivation to learn by presenting training that is engaging.

Use of Assessments

With increasing expectations for learners to guide their own learning, the Army needs to develop strategies and employ technological tools that foster self-directed learner investigation. Accurate methods of tracking student progress (e.g., learning, progress against learning goals) will need to be matured and employed if the benefits of the dynamic training system proposed by the ALM are to be realized (i.e., individualized learner-centered ubiquitous training).

Specifically, valid and reliable assessment will be essential in future Army learning practices. Assessments will be necessary to ensure that learning has occurred to a standard, and frequent assessment will be necessary to track learner's progress and tailor instruction to support a truly adaptive learning system. The frequency of assessment to track learner progress can be viewed over a time continuum, and can be as simple as a pre- and post-test or can include progress tests, which assess changes in learner progress at various points through the training cycle. Progress tests allow for greater tailoring of instruction during the training. Similarly, assessment can occur within a single platform (i.e., method of training delivery) and as such is referred to as within-platform assessment, or can occur across an entire training effort, including several or all platforms (i.e., cross-platform assessment). Regardless of the type, assessment is a key leverage point for enabling technology to sustain a learner-centered operational training environment.

Current Research

While the ALM recognizes the important roles of both learning technology and assessment for realizing and sustaining its desired learning practices, more research is needed to inform what specific technology and assessment approaches should be employed and how and under what conditions they should be implemented. Specifically, the Army needs guidance on how to use and integrate assessments and various learning technologies in a way that will meet the goals of the ALM.

To address this need, the U.S. Army Research Institute (ARI), Orlando is collaborating with the Army Research Laboratory Human Research & Engineering Directorate Simulation & Training Technology Center (ARL-HRED-STTC) on a project to develop prototype training materials and assessments to test ALM concepts in an integrated, technology-enabled environment. Under this project, which is known as the Soldier-Centered Army Learning Environment (SCALE) project, ARI has contracted with ICF International (ICF) to develop and test a prototype of Army training that is aligned with the ALM. Specifically, this prototype training is to follow the approach presented in Figure 1.

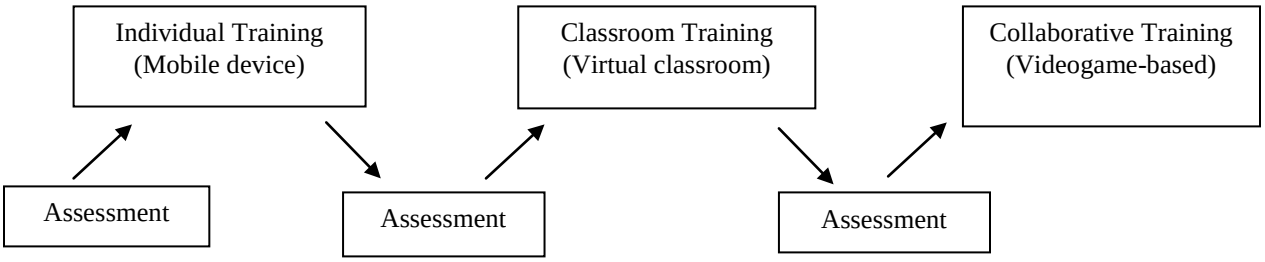

***Figure 1*:** SCALE prototype training diagram.

The goal in the development and testing of this prototype training is to investigate different approaches to training assessment development and delivery using emerging technology as specified in the ALM. Many forms of assessment exist and can be used throughout training to assess learning and progress. Each form of assessment contains characteristics that are advantageous (or may limit their use). Computer-adaptive tests, which adapt test content to the test-taker's ability through computer algorithms, achieve accuracy with fewer test items (Mead & Drasgow, 1993) and have been successfully demonstrated on mobile devices (Triantafillou, Georgiadou, & Economides, 2008). Although they typically require significant dedication of resources to develop, the precision that these types of tests can provide makes adaptive tests very relevant to the goals of the ALM.

By investigating different approaches for developing the prototype training, ARI seeks to answer six key research questions that address the use of assessments and learning technologies:

- How should assessments be designed, delivered, and otherwise used to maximize Soldier training?
- How should adaptive assessments be implemented?
- How often should assessments be conducted?
- What are Soldiers' preferences for training on technology-based platforms?
- How effective is training that is delivered through technology-based platforms?
- What are best practices for delivering and developing training evaluations to maximize the benefits of leveraging these emerging technologies?

As an initial step to answer some of these research questions and help guide the development of the prototype training, we sought best practices for developing and delivering assessments on mobile devices and virtual platforms. Specifically, ARI aimed to find exemplars within the U.S. Army, other U.S. Armed Services, or other nations' Armed Forces, and the private sector in

which training and assessment were carried out operationally using the following platforms: mobile devices, virtual worlds, and videogame-based scenarios, which align with the prescribed approach for the SCALE prototype training. From these exemplars, best practices were identified for integrating mobile, virtual and videogame-based training, address the role of assessments and the extent to which they can be administered and managed through these different platforms, discuss the benefits and challenges involved of using these technologies, and provide guidelines for future training and assessment development.

Identifying Exemplars

Our strategy to identify exemplars and best practices around the use of assessments in the platforms of interest (mobile devices, virtual worlds, and videogame-based scenarios) was to conduct a thorough literature review and interviews with training and assessment experts. To be informative for the SCALE prototype training, the exemplars needed to meet certain criteria. Of highest relevance, as identified in the research requirement, were training activities that integrate multiple platforms. Added interest was on those training efforts that exemplified a best practice when incorporating assessments that offered multiple levels of feedback to stakeholders such as learners, instructors, course designers, and unit leadership. The research requirement also stipulated that the training needed be operational and occur within the U.S. Army, other U.S. Armed Services, other nations' armed forces, or the private sector. Research reported on K-12 populations was included only if the assessment strategy could generalize to Soldier training or if there was an integration of the platforms of interest. In addition to these criteria, we determined that the training needed to represent a single training effort and have empirical evidence in order to support the identification of best practices.

These qualification criteria are outlined in Table 1 and are listed in order of importance. The third criterion is a measure of relevance based on how many platforms of interest were included, and empirical strength based on 27 factors, such as whether there was random assignment of subjects, whether the reliability of scales was reported and was acceptable, and whether there were multiple post-tests. This measure was developed by the research team to ease comparison across sources. Reported trainings were reviewed to determine if they met all of the criteria to be considered an exemplar.

Table 1

Qualification Criteria for Exemplars

Criterion	Description
1. Distributes training via two or more platforms in the following categories: mobile device, virtual worlds, or videogame-based scenarios	Training activities included training across multiple platforms, preferably all three platforms of interest.
2. Includes individual and/or collective assessment of learning progress and/or assessment of one or more learning outcomes	The training focused on methodologies to assess and evaluate training, with attention paid to the level of evaluation (i.e., reactions, learning, behavior, outcomes). Statistical analysis of training effectiveness included at least a within-platform assessment of pre- and post-learning.
3. Is relevant and empirically strong	Research on the training was empirically strong and relevant to the SCALE prototype training based on an ad hoc scale developed by the research team.
4. Is used in an operational setting	The training has been implemented in applied, operational or field settings with real learners who receive credit or recognition for training completion. Not a 'proof-of-concept', test product, or prototype that has yet to reach an operational stage.
5. Is a discrete unit of training	The training represents a single, discrete unit or course of training as opposed to an entire training program curriculum.
6. Implemented within or for groups of interest	The training occurred within U.S. Army, other U.S. military Services, other nation's armed forces (e.g., NATO), government personnel, or private sector.

Knowing that the chance of finding an exemplar that met all of the criteria would be low, and given the current state of the research in this area, a critical component of our strategy was to identify "exemplary elements" from sources. Exemplary elements refer to singular training practices, assessment events, or clever applications that were judged to be of a high standard, but were not necessarily integrated across the platforms, or did not necessarily assess learner performance across a time continuum. These elements could serve in the formulation of a synthesized ideal in the absence of all-inclusive examples. Examples of exemplary elements include highly effective training deployed in a single mode, or educational programs in classroom settings that provide innovative development or implementation of an assessment or evaluation process. Thus, the technical approach was to seek full exemplars while recognizing

exemplary elements in order to capture as much information as is currently available to inform the development of best practices and guidelines.

Method

Literature Review

Sources. A literature search of online databases (Education Resources Information Center, PsycINFO, PsycARTICLES, PsycEXTRA, Psychology and Behavioral Sciences Collection, and Defense Technical Information Center) and electronically available conference proceedings was conducted using the search terms listed in the first column of Table 2. Given the breadth of the Google Scholar database, a Boolean strategy that was used combined terms in the first column with the operator, AND, with terms in the second column of Table 2 to guide the search. Literature sources included, but were not limited to, technical reports, peer-reviewed professional journals, military journals, periodicals, book chapters, and presentations from relevant defense and/or private-industry conferences and seminars. Approximately 1,200 sources were found using these databases and search terms.

Table 2
Search Terms for Literature Review

Search Term 1	**Search Term 2**
mobile learning	experimental
m-learning	control group
ubiquitous learning	evaluation
u-learning	evaluate
learner-centric	assessment
personalized learning	pre-test
adaptive learning	post-test
blended learning	performance
hybrid learning	effectiveness
e-learning	utility
dL	outcome
distributed learning	cross platform
virtual classroom	mixed modes
virtual training	
online training	
computer-based training	
collaborative learning	
simulation	
multiplayer simulation	
serious game	

Table 2 ***(continued)***

Search Term 1	Search Term 2
simulation-based training	
virtual world training	
computer-based simulation	

Note. Italic search terms also substituted 'training' with 'learning' or vice versa. Search Term 1 and Search Term 2 variables were combined when searching Google Scholar. For example, mobile learning was searched using all Search Term 2 options (ex. mobile learning+experimental, mobile learning+control group, etc.).

Procedure. The abstracts of the identified sources were scanned to determine if the source provided an evaluation of a training that included one or more of the platforms of interest, implementation or development of an assessment or training evaluation process, or discussed lessons learned regarding the platforms of interest or assessments. If the abstract did not provide enough information, the entire source was scanned for these elements. Seventy-seven sources included one or more of these elements and were retained for full coding.

Three of the authors served as coders and coded the 77 sources to capture information regarding the type of training platform used (e.g., mobile, videogame, virtual world, virtual classroom, traditional classroom), the type of content that was trained (e.g., procedural, cognitive, affective), the software that was used (e.g., COTS, other), the type of assessment that was included in the training (e.g., platform-specific, cross-platform), the level of outcomes that were measured during and/or after the training (e.g., reactions, learning, behavior, organizational), the type of research design that was used to evaluate the training (e.g., within subject, between subject, mixed), how the training was evaluated, and other information pertaining to the source's research methodology (e.g., sample size, reliability, effect sizes). A Microsoft© Access database was used to capture the pertinent information and included a mix of yes/no and text/numerical fields. Appendix A provides a list of the final fields included within the database along with a short description of each field (where applicable).

Analysis. The information contained in the coding database was used to identify exemplars and exemplary elements. As such, it was important to ensure sources were coded in a consistent manner. Coders were part of the database development team and had multiple team discussions regarding the database fields. Two sources were then randomly selected and coded independently by all three coders. Consistency was assessed through a group meeting in which the three coders discussed their data entries for the two sources to determine the percent of agreement. Coders demonstrated 100 percent agreement on all non-text fields. Discussion of coders' text-based fields demonstrated that coders shared a similar interpretation of the articles. Differences in text-based fields were due to differences in the level of detail that coders provided, not in the meaning of the response. If coding questions arose during the remainder of the coding process, coders discussed the question until a consensus was reached.

Next, the information in the database was reviewed for each of the 77 sources to determine which sources could be considered exemplars. To be considered an exemplar, the training reported in the source had to meet all of the criteria provided in Table 1 (i.e., Distribute training via two or more platforms of interest, include individual or collective assessment of learning, be relevant to the SCALE prototype training and have empirical strength, be

operational, encompass a discrete unit of training, and be developed for one of the groups of interest). The third criterion regarding relevance and empirical strength was a scale created by the research team to quantitatively summarize the methodological rigor of the research provided on the reported training and the degree to which the reported training was relevant to the SCALE prototype training (i.e., how many of the platforms of interest the training incorporated).

Different point values were assigned to reported trainings depending on the number of platforms of interest that were included (Table 3), the level of outcomes and type of assessments used (Table 4), the type of experimental design that was involved (Table 5), and the features of the evaluation approach used (Table 6). Trainings could receive multiple point values for each factor if they included more than one of the features listed. For example, if a training included 2 platforms of interest (2 points), a multiple choice level 2 assessment (1 point), a situational judgment level 2 assessment (2 points), a team-based simulation (2 points); was evaluated through a pilot project (1 point) that used a between subjects design (1 point) with a pre- (1 point) and post-test (1 point); and reported descriptives (1 point) and effect sizes (1 point), the training would receive a score of 13. Scores were calculated for each of the 77 sources after they were entered into the database. The range of scores achieved across the 77 sources was 0 to 19 points.

Table 3
Relevance and Empirical Strength Scale Points for Number of Platforms Factor

Number of Platforms of Interest	Points Awarded
1 platform of interest	1 point
2 platforms of interest	2 points
3 platforms of interest	4 points

Table 4
Relevance and Empirical Strength Scale Points for Assessment Level and Type Factor

Assessment Level	Assessment Type	Points Awarded
Kirkpatrick Level 2	Unknown or unclear	1 point
	Multiple choice test	1 point
	Situational judgment test	2 points
	Adaptive test with adaptive instruction	3 points
Kirkpatrick Level 3	Unknown or unclear	2 points
	Individual performance in a simulation or exercise	2 points
	Team-based performance in a simulation or exercise	2 points
Kirkpatrick Level 4	Organizational results	1 point
Not applicable	Attrition	1 point
	Knowledge or skill retention	1 point

Note: If a training included assessments at multiple levels and/or multiple types, it received points for each assessment.

Table 5

Relevance and Empirical Strength Scale Points for Experimental Design Factor

Experimental Design Features	Points Awarded
Quasi-experimental	1 point
Comparison group	2 points
Between subjects	1 point
Within subjects	2 points
Mixed subjects	3 points
Matched sample	2 points
Pilot project	1 point
Meta-analysis	1 point
Case study	2 points

Note: If a training included multiple experimental design features, it received points for each feature. Points awarded were derived according to strength and complexity of the experimental design.

Table 6

Relevance and Empirical Strength Scale Points for Evaluation Factor

Evaluation Features	Points Awarded
Descriptives Reported (Yes/No)	1 point
Testing Frequency - Pre-test	1 point
Testing Frequency - Single Post-test	1 point
Scale Reliability Reported (Yes/No)	1 point
Effect Size Reported (Yes/No)	1 point
Randomly Assigned Treatment (Yes/No)	2 points
Testing Frequency - Progress-test(s)	2 points
Testing Frequency - Multiple Post-tests	2 points

Note: If a training included multiple evaluation features, it received points for each feature. Points awarded were derived according to strength and complexity of the experimental design.

Next, reported trainings were reviewed that did not meet all of the criteria in Table 1 to identify best practice data in the form of exemplary elements. Again, exemplary elements refer to singular training practices, assessment events, or clever applications that were judged to be of a high standard and represent a unique or innovative approach, but did not meet the criterion of an exemplar. These elements provided insightful approaches and recommendations on how to assess performance or integrate assessments within the platforms of interest.

For each exemplar and each exemplary element that was identified, a “key takeaway” statement was written describing the practice inferred by the exemplar or elementary element and how it related to the ALM.

Interviews

Participants. Twenty-two subject matter experts (SMEs) were interviewed that had been identified through literature search activities, review of professional associations, and recommendations from other SMEs. To be considered as a potential interview SME, the individual had to meet one of the following criteria: 1) have specific knowledge of assessment methods within learning technology platforms, 2) be a leading researcher or academician with a known record of research and publication within the domains of training and assessment (i.e., publication productivity in refereed journals), 3) have extensive knowledge of training implementation across relevant learning technology platforms, or 4) have knowledge of specific potential exemplars. The participating SMEs included five academicians, five researchers, five private consultants, four training developers, and three training program managers. Of the 22 SMEs interviewed, nine represented private sector organizations, four represented U.S. defense organizations, six were from universities, one represented a non-profit association, one was from a defense organization of another nation, and one represented a U.S. federal agency.

Protocols. Two interview protocols were developed that used a blend of structured inquiry to ensure consistent data were gathered on the central research questions, and unstructured inquiry to gather relevant information based on the unique experiences of each SME. One protocol was created for training developers with the expectation that they could provide detailed information about specific assessment development steps. A second protocol was created for training managers and researchers, who were assumed to be able to provide information regarding policies, intentions for training (e.g., underlying organizational goals), and higher-level issues associated with assessment. The two protocols were highly similar and included questions regarding SME and training program background information, training platforms, development of training activities, assessment within training, training evaluation, and results and future plans. Both protocols are in Appendix B.

Procedure. Interviews were scheduled and conducted by telephone using the appropriate protocol based on the type of SME being interviewed. Two research team members, a facilitator, and a note taker – typically conducted the interviews, which generally lasted 30 to 60 minutes. At the end of each interview, SMEs were asked about additional exemplars they may be aware of, either within their organization, or outside of it (e.g., through their own research or participation/contacts in the training community), and thanked them for their participation.

Analysis. The notes from each interview were reviewed and analyzed for key comments that provided insight about the implementation or development of an assessment or evaluation process, or the integration or use of the platforms of interest. These comments were referred to as "interview insights." A "key takeaway" statement was written for each interview insight describing the practice inferred by the insight and how it related to the ALM.

Synthesis of Findings

The exemplars, exemplary elements, and interview insights, including their key takeaways, were analyzed to identify discernible patterns or consistent themes within the

following categories: 1) best practices for integrating mobile, virtual and videogame-based platforms, 2) role of assessments and how they can be implemented within these platforms, or 3) benefits or challenges of the platforms. For the purposes of this research, a theme was defined as a practice, recommendation, or idea that either was identified more than once within or across data collection sources, or was particularly innovative and/or very relevant to the goals of the ALM. Themes were initially identified by research team members individually. Then, a group meeting was held to discuss all of the suggested themes. Only themes that received consensus agreement were retained.

A set of provisional guidelines and proposed practices were developed in accord with the themes in each category. Given the infancy of the research in this area, there are no industry standards or body of evidence to support particular activities around the use or integration of assessments within the platforms of interest; therefore the practices and guidelines can only be considered proposed or provisional. The overall methodology and how it led to the provisional guidelines and proposed practices are outlined in Figure 2.

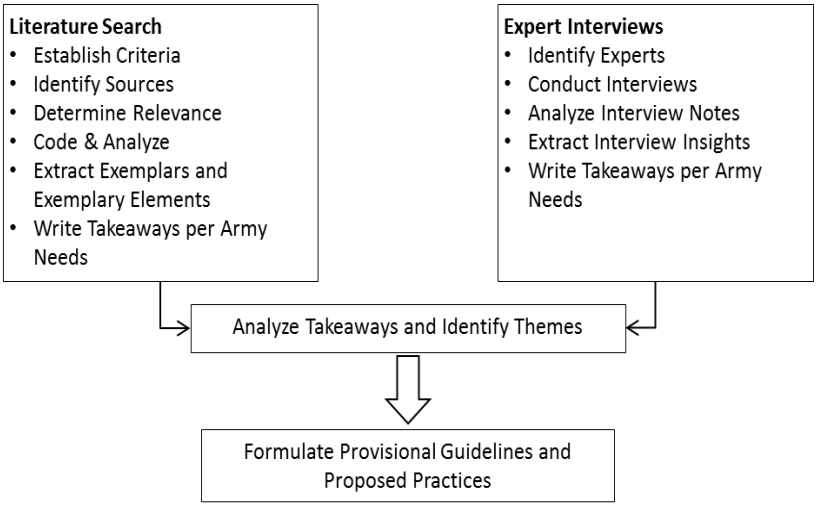

***Figure 2*:** Research methodology overview.

Results

Exemplars

From the 77 sources, 3 exemplars and 23 exemplary elements were found (See Appendix C). While over half (55%) of the 77 sources described distributed training via one or more of the platforms of interest and 83 percent of the sources incorporated some type of individual or collective assessment, very few (17%) were implemented in an operational setting. The mean score on the Relevance and Empirical Strength criterion across the 77 sources was 9.18 (SD = 4.95, range = 0-19 points), indicating that most sources did not include multiple platforms and lacked the empirical rigor to be confident of the results. The 23 exemplary elements that were identified described unique or innovative approaches that represented an effective practice. The remaining sources did not meet the criteria for exemplars and did not provide an approach to

integrating the platforms of interest or to the use of assessments that could be considered unique, innovative, and/or an effective practice.

Of the three exemplars, two were developed for and delivered to U.S. military audiences, while the third was developed for college students. None of the exemplars utilized all three of the training platforms of interest, but each did employ multiple training platforms, with at least one or more of those platforms being a mobile, virtual world or videogame platform. The content taught in all three of the exemplars was cognitive skills, with one of the exemplars also including procedural and psychomotor skills. All three exemplars included a pre-test and one or more post-tests. Two of the exemplars also included progress tests (i.e., tests administered during training to assess changes longitudinally). All three exemplars assessed two or more levels of outcomes, with all three including measures of learning outcomes, which were either multiple-choice or situational judgment tests. None of the exemplars utilized adaptive testing to assess learning outcomes.

Exemplar 1 – Montijo, Spiker, & Nullmeyer (2010). This exemplar involved multi-platforms to train crew who fly remotely piloted aircraft called Predators. The goal of the training was to reduce mishap-related errors and focused on improving Task Prioritization, Channelized Attention, Selection of Appropriate Course of Action, and Crew Coordination. The training involved the use of four platforms: 1) facilitated classroom training, 2) computer-based training providing information on case histories, 3) a videogame for students to practice individual skills, and 4) a simulation game that practices crew coordination in a stressful environment.

The research methodology was structured around the concept of 'spirals' (i.e., differing sets of the training platforms) that were administered to different samples of students. For each spiral, researchers collected data on reaction and learning outcomes (i.e., Kirkpatrick levels I and II), in addition to two cross-platform assessments (i.e., simulated exercise and a flying mission), which were conducted at the end of the training to assess behavioral outcomes. Organizational outcomes (i.e., impact) were collected through a survey of supervisors.

Evaluation of the training was conducted by comparing results of learning and behavioral assessments among the various 'spirals.' Researchers used hypothesis testing to determine if there were significant differences in the results of both learning and behavior. Results showed that the spiral including all four training platforms showed a significantly positive difference for learning over the first spiral (i.e., facilitated classroom only) and the second spiral (i.e., facilitated classroom and computer-based training), while comparisons with other spirals were not significant. When comparing results of the behavioral assessment, the spiral including all platforms also showed significant improvement over the spiral containing only classroom training. No other comparisons showed statistical significance.

In summary, the training effort demonstrated effective practice in using multiple learning technology platforms incorporated into a single training. It also aligned content type with the most appropriate training platform (i.e., declarative knowledge through classroom and computer-based training, behavioral competencies through an interactive video-game platform, collaborative decision-making and social skills through a collaborative virtual simulation). In

addition, the effort demonstrated effective practice in the use of assessments by utilizing within-platform and cross-platform assessments. Finally, the effort provided an example of a well-planned, comprehensive and innovative evaluation effort within this context.

Exemplar 2 – Ross & Kobus (2011). The goal of the training in this exemplar was to improve decision-making within small dismounted infantry units under a variety of conditions using a high-fidelity immersive virtual simulation i.e., the Future Immersive Training Environment (FITE). The FITE is a decision-skills trainer that immerses learners into various scenarios and provides mixed reality experiential learning. It encompasses an integrated suite of technologies, such as intelligent avatars, natural language interfaces, animatronics, instrumentation, and replay capabilities. Training in the FITE platform was partnered with classroom training in a 'blended learning' approach.[1]

The assessment strategy for the training included a situational judgment test (SJT), which was based on cognitive task analysis (CTA). SJTs pose realistic yet hypothetical problems to test-takers and ask them to provide an appropriate response (Christian, Edwards, & Bradley, 2010). The SJT assessment was composed of a number of vignettes written to assess 27 'decision themes' that were the output of the CTA. Two vignettes were developed for each theme, and each vignette was developed into an SJT item by collecting feedback from SMEs. For each pair of SJT items, one was assigned to a pre-test assessment and the second was assigned to a post-test assessment. The SJTs were partnered with a more qualitative after action review (AAR). Results from the pre-test and post-test SJT assessments were targeted at measuring decision-making in the individual team member, while the AAR was focused on providing feedback on team-level decision-making and behaviors.

Evaluation of pre- and post-test results indicated significant improvement in individual and team decision making across the identified themes. This activity was exemplary in its use of assessments in a blended approach to training. Specifically, the approach of combining an assessment to measure individual competencies (SJT), partnered with an additional assessment to critique team-based performance (AAR) was a unique strategy to assess the outcome of the training, particularly when applied to the results of team-oriented scenario-based virtual simulation.

Exemplar 3 – Chen, Chang & Wang (2008). In this exemplar, a freshman-level university computer science course created a ubiquitous learning environment that integrated a variety of training platforms to include mobile learning, virtual classroom, computer-based training (CBT), and traditional classroom lectures. The virtual classroom condition focused on providing peer-based mentoring delivered via mobile phones, personal digital assistants (PDAs) or computers using chat and messaging functions.

Learner progress was assessed in several ways. First, learning was assessed through quizzes and multiple-choice assessments delivered online throughout the training course.

[1] While this training did not include more than one of the platforms of interests, it was still included as an exemplar, as determined by the research team and client, because it demonstrated an effective marriage of the advanced learning technology with the traditional classroom and best practices in situational judgment test development.

Second, the length of time spent interacting with the content was measured for each learner. Third, learners completed a self-assessment of learning progress. These three measures were combined with individual learning goals and information on student schedules to create a student model that adapted delivery of training content to the learner (i.e., assigned learning tasks). A key component of the system was the incorporation of reminders and alerts (e.g., text messages) delivered to students to inform them of their learning status and reminding them of learning goals relative to the course timeline. This component was the focus of the evaluation and the results showed that those students receiving the 'awareness tools' (i.e., text-based reminders) scored higher than students using the system that did not receive the reminders.

This training demonstrated integration of multiple learning technology platforms to create a ubiquitous learning environment, tracking of learner progress across multiple training platforms (throughout a training course), and the use of collected learner data to adapt training delivery. The course also highlighted the utility of automated tools that promote learner awareness of progress, and how this awareness can affect learning outcomes.

In the next sections, themes that were identified from the analysis of the exemplars are presented, along with the exemplary elements, and interview insights.

Themes Regarding Best Practices for Integrating Platforms

The integration of training platforms is an area that is just beginning to be researched: only 15 of the 77 sources in the literature review used or compared more than one of the platforms of interest. The following themes were identified regarding the integration of mobile, virtual, and videogame-based scenarios from an evaluation of the three exemplars, the exemplary elements and interview insights.

Importance of learning sciences' principles. Although innovative training technologies are widening the possibilities of training design and delivery, the findings consistently emphasized that basic training principles and tenets from the learning sciences should not be forsaken and should be applied in the development and implementation of learning frameworks. Effective integration requires careful planning. Regardless of training platform, effective strategies, such as those related to instructional systems design (ISD) should be used to carefully plan the training development and implementation process. Insights from the interviews suggested using ISD principles as a way to make decisions about 1) training implementation (e.g., delivery environment/technology), 2) programming of learning objectives, and 3) incorporation of content. Furthermore, when the appropriate learning science strategies are followed, training design patterns can then be shared between games and between virtual classroom tools (Mautone, Spiker, Karp, & Conkey, 2010; Salmon, Ming & Palitha, 2010).

Part of careful planning for training and integration should involve the employment of a strong group of SMEs that includes content SMEs as well as training designers and technology-platform developers (Okuda, Arcaro, & Gaught, 2011). These SMEs should be involved throughout the development process.

Transferability of training content. As learned through the interviews, consideration should be given to how training content can be effectively shared among training platforms during the planning stage. Having the ability to migrate training content from one platform to another, although not necessary, can be a key benefit. During the interviews, one SME provided a salient example involving a training development project for which content was custom-developed for integration in a virtual-world training platform. Following development of the virtual-world training, members of the organization expressed an interest in delivering the training content through another platform. Because there was no initial forethought into sharing the content among delivery methods, there was a high degree of difficulty 'extracting' the content that had been developed in one type of software for use in another type of platform. However, if the same content is available on multiple platforms, the learner is afforded with greater training options.

Use of tools for training platform decisions. Determining if a specific platform is optimal for training can be a difficult task. Creating and utilizing decision tools for such determinations is possible and useful. For example, a rule-based decision tool was used in one of the exemplary elements to determine which training areas would benefit from game-based training (Mautone et al., 2010). These tools can help ensure decisions about the use and selection of technology platforms are driven by the learning objectives and not just a desire to use innovative technology.

Alignment of learning objectives to environmental context. In addition to selecting appropriate training platforms for the learning objectives, one of the exemplary elements pointed to the importance of aligning the learning objectives to the most relevant context within the appropriate training platform (Tichon, 2007). For example, the rule-based tool mentioned previously also makes recommendations about which game elements and design patterns within a game-based platform would be most appropriate for the learning objectives. It is important to understand the ability of specific platforms to create situated learning opportunities and to align those so that they are most similar to the environmental context in which the skill will be transferred and implemented. As an example, using GPS enabled devices for field training related to mapping and reconnaissance skills is preferred, from a situated learning perspective, over traditional paper and pencil instruction. Aligning learning objectives to relevant environmental contexts allows for easy transfer of trained knowledge to an operational environment.

Themes Regarding the Role of Assessments

Future Army training goals repeatedly reference the importance of assessment to improve Soldier training. Accurate and reliable assessments are essential for an adaptive training environment that 1) tailors training to meet individual Soldiers' needs and levels of proficiency, 2) provides individualized feedback, 3) tracks Solider performance within and across platforms, and 4) validly measures a performance criterion.

According to the literature review, the use of assessments within the platforms of interest are common, with half or more of the sources for a given platform including some form of assessment at some point. However, what was notable was the limited use of progress tests in

the platforms of interest within the research literature. Progress tests were included in less than 20 percent of the sources that utilized mobile platforms, approximately one-third of the sources for videogames and virtual worlds, and about half of the sources that utilized virtual classrooms. Progress tests are necessary if truly adaptive training is the goal; without initial and multiple assessments, training content cannot be tailored to an individual trainee's current knowledge or proficiency level. This oversight is a large setback considering that assessment of performance in virtual environments is required for tailoring instruction; assigning levels of competency or proficiency to learners, tracking trainee performance, and evaluating overall training system effectiveness (Pokorny, Haynes, & Gott, 2010), all of which are goals of future Army training. Still, from the exemplars, exemplary elements and interview insights, the following themes were identified that related to the role of assessments and how to design and/or implement them within the platforms of interest. From these themes, provisional guidelines and proposed practices were developed.

Importance of planning for assessment. Just as training design teams must rely on ISD principles to design effective training, design teams must also apply guidelines and principles to plan assessment strategies. Specifically, it is important to plan the level, scope, detail, and impact of assessments in areas such as 1) determining the knowledge, skills, and abilities (KSAs) of interest for the assessment, 2) determining the behaviors and performance indicators of these KSAs, and 3) determining what actions would indicate that learning had occurred. As noted from the interviews, if the behaviors indicative of the KSA are unclear, assessment is a moot point. For example, if trainers are interested in assessing leadership emergence but fail to define and operationalize this competency, or the indicative behaviors, assessing the level of leadership emergence that learners demonstrate would prove problematic.

An assessment map is one way to begin framing an assessment strategy. Event-Based Approaches to Training (EBAT) and Evidence-Centered Design are two examples of assessment mapping techniques (Rosen, Salas, Weaver, Lazzara, King, & Robinson, 2010). Both approaches emphasize determining the linkages between the competencies of interest, the indicators of these competencies within a training environment, and the assessment options and techniques. These decisions are best made by teams of experts in the subject matter of interest as well as in performance assessment (Okuda et al., 2011). As an example, the use of EBAT would include collecting critical incidents related to competencies of interest and engineering these competencies to assessment components. These components can then inform scenario design so that there are opportunities to demonstrate the competency-linked KSA (Rosen et al., 2010).

Another important insight from one of the interviews was the potential for data overload. Technologically innovative platforms supply a mass of data that can become unwieldy to interpret if assessment decisions are not made a priori. Thus, careful consideration should be given to training content and technology features when creating an assessment plan to determine what information will be used and how it will be evaluated.

Finally, assessment planning can illuminate potential issues that may arise from the type of platform used to administer the assessment. For example, as noted in one of the interviews, comparing speeded test scores on mobile phones to a desktop computer may be more indicative

of the display and keyboard size differences and less demonstrative of differences in the KSA. Thus, the platform could potentially become an unexpected confound in the assessment process.

Use of one platform for both training and assessment. The exemplars, exemplary elements and the interview insights provide numerous examples of using the training platform to deliver training and assess the learner. Such practice would address future Army goals that specify the need for real-time assessments. For instance, relevant examples assembled from the qualitative data included:

- Using videogames as an assessment tool (McDowell, Johnson, Freeman, Roberts, & Horn, 2011),
- Building calculators in simulations to assess performance,
- Making assessments a structural component within a game or simulation (Bowling, Khasawneh, Kaewkukekool, Jiang, & Gramopadhye, 2008),
- Using knowledge checks throughout training and linking the knowledge checks to work actions,
- Streamlining the observation process through mobile devices or electronic checklists,
- Embedding work samples into a simulation or game and scoring virtually by expert judges (Ross & Kobus, 2011),
- Assessing content of communications in asynchronous discussions and chat rooms (So, 2009; Wang, Newlin, & Tucker, 2001), and
- Monitoring learner states in an intelligent mobile learning system (Chen & Hsu, 2008).

Platforms such as videogames have the ability to capture trainee data and calculate performance scores based on game play. For example, if training occurs in a videogame-based environment and videogame performance is indicative of the criterion of interest, the videogame is providing assessment, as well as training. Insights from the interviews revealed promising uses of testing-as-training such as:

- Testing as a form of learning
- Testing in continuous assessments by mobile devices
- Testing that is invisible to the user as a structured learning component (i.e., stealth assessment)
- The application of scoring rules in creating work samples
- Testing delivered while engaging in video-game based learning scenarios
- Testing of immersion-related behaviors used as surrogate assessment measures in virtual worlds

The interviews suggested that because assessments can both train and assess, more assessments equate to more training. Evidence from the literature illustrates uses of this practice by engineering content into assessment components (Rosen et al., 2010). Utilizing the capabilities of the technology to assess the learner while the learner is engaged in the training

capitalizes on the efficiencies that can be gained through the technology (e.g., single platform, electronic data capture), assesses the learner in the way the learner has been trained, and minimizes disruption to the learner.

Automated assessment through the same platform as the training also enables machine-based decisions about training progress to allow for the tailoring of training to match Soldiers' needs. However, assessment design should take into account the nature of the incorporated training content (e.g., procedural, knowledge, psychomotor, affective), a Soldier's prior level of knowledge, and the complexity of the environment (i.e., the need for a human-in-the-loop testing scheme) to determine if automated assessment is feasible. Although there are those who claim that the potential exists to train and assess almost any knowledge, skill, or ability in a mobile or virtual environment, there was still consensus from the interview SMEs that highly complex training content is not yet suited for automated assessment.

Importance of adaptive instruction. While emphasizing the need to train Soldiers to be adaptable, the ALM also stresses that training content and the way it is delivered must be adaptable or adjustable to meet the needs of Soldiers. The model articulates that adjustments to training should occur based on operational changes, Soldier performance, and advances in training technologies. In addition to those changes, to support a learner-centric environment, training should also be adjusted based on the learner's prior experience, individual differences, and/or current performance. Adaptive instruction, a common term for training that is adjusted based on the learner, was reported to be used in reading instruction (Chen & Hsu, 2008), kinematic instruction, and e-learning packages embedded as diagnostic tools (Kalyuga, 2006a), all using the platforms of interest.

The interviews showed many ways assessments can be used to support adaptive instruction, such as using results to adapt training content to an individual learning style (Tucker & Goodwin, 2010), or using assessment of behavioral performance in a virtual environment to integrate strategies for adapting learners to virtual world content.

Use of adaptive assessments. Assessments, themselves, can also possess adaptive capabilities. For example, assessments can increase or decrease in difficulty or focus on a particular content area based on learner responses. One limitation to using adaptive testing, as pointed out in the interviews, is that some complex stimuli and constructed responses may not be positioned to take advantage of computer adaptive testing due to the difficulty of calibrating non-multiple-choice items. Further, the interview insights noted that the development of items is time consuming and requires some level of expertise. These drawbacks may be the reason that only one of the 77 sources used adaptive testing. Both adaptive training and adaptive assessments create optimal training environments by providing tailored training to learners; however, both necessitate considerable planning and investment.

Importance of frequent testing. Examining pre- or post-tests in isolation limit the ability to make valid conclusions about the success of training (i.e., improvement in some competency of knowledge base). For example, inferences about the relationship between training and a post-test of declarative knowledge are greatly limited if a pre-test of declarative knowledge is not included to assess the amount of change in declarative knowledge. Including

both pre- and post-tests increase the strength of training effectiveness claims; however, more frequent testing (i.e., progress tests) allows for training to be assessed as a process as opposed to a snapshot in time, as noted in the exemplars, exemplary elements and interviews (Chen et al., 2008; Guzman, Conejo, & Perez-de-la-Cruz, 2007; Lancaster & McQueeney, 2011). Further, continually adaptive training necessitates frequent testing. As noted in one of the interviews, current performance levels are needed to determine the training or the training path that should be delivered next.

Utilization and evaluation of social exchanges. The ALM highlights the importance of using peer-based and collaborative learning as a strategy for cultivating a learner-centric environment. Evidence from the interviews provided specific examples of how this strategy can be implemented, such as developing communities of practice where geographically dispersed individuals can collaborate and establish unstructured learning networks, the development of "wiki" technologies to assist in co-creation of knowledge among peers, and fielding mobile applications that encourage users to modify the application for their own use and to provide specific functional suggestions to training developers to make the content more relevant to peers.

In addition, social exchanges between learners within a platform may be used for assessment if these exchanges are linked to competencies. Interview insights and an exemplary element noted the potential benefit of collecting and analyzing data generated from social exchanges (e.g., e-mails, text messages, dialogue in on-line discussion forums) in collaborative environments mediated by technology (So, 2009). Specifically, one interview insight suggested that data from social exchanges could be combined and analyzed across exchange mediums and among all parties participating (i.e., learners, collaborators, instructors) to obtain the most informed picture of how peer and collaborative learning is taking place and how learners with different learning styles, such as inductive versus deductive preferences, are responding to the training. Social exchange behavior, if operationalized in such a way, also has the ability to be indicative of competencies such as teamwork, collaboration, leadership, etc.

The literature review provided specific examples of how social exchange data has been collected and used in specific training platforms. Examples showed mapping of social interactions within virtual worlds, and analysis of data from online discussion forums within virtual classrooms (So, 2009). In addition to course assignment grades in virtual classrooms, other sources analyzed social exchanges, such as chat-room conversations or message board posts, within virtual classroom environments as a means of assessing trainee collaboration or various learning outcomes (typically via content analysis). These assessments were mostly used to inform instructors/trainers/designers on how learners used the technology, factors that influenced responding through the medium, and experiences by learners that affected their perceptions of the utility of the training tool (So, 2009). Discourse analysis, as discussed by Wang et al. (2001), is but one way to analyze the content of chat-room activity. Discourse analysis encompasses the use of chat log data as well as rater coding to assess the quality of chat-room based content, in addition to the quantity of chat room activity. The key is to plan for what aspects of social exchange data are relevant to the learning objectives.

Themes Regarding Benefits and Challenges of Platforms

Benefits and challenges of mobile devices. Research on mobile learning (e.g., Holden & Sykes, 2011; Norris & Soloway, 2004; Rochelle & Pea, 2002; Soloway, Norris, Blumenfeld, Fishman, Krajcik, & Marx, 2001) cites a number of potential advantages for mobile learning technology. First, mobile devices give learners flexible access to learning materials, which can increase learning gains. Within an Army context, Tucker (2010) cites benefits such as more engagement with peers, more collaborative relationships between learner and instructor, and greater time on task due to the motivating features of mobile devices. Second, mobile devices now offer much more than merely "e-learning with mobile devices" (Ally, 2004 – cited in Frohberg, Goth, & Schwabe, 2009). By shifting training devices into the hands of Soldiers, mobile devices place learning in the context and time demands of the real world, which according to situated cognition theory should be beneficial (Lave, 1988). Mobile devices can create situational learning opportunities by using features such as global positioning system (GPS) or various communication capabilities (including voice, text, photographs, and audio recordings) that allow collaboration between instructors and peers similar to real world interactions (Hwang & Chang, 2011). Such features can be applied in instructionally relevant ways to mimic operational tasks such as urban combat, reconnaissance and selection of position, or interrogation.

The challenges to mobile learning often center on usability issues, such as trying to view learning content on a small screen. When details are important, the smaller screen can become a limiting factor. There are also issues with security concerns (e.g., potential theft), signal strength in field settings, and the ability to provide extended text feedback to instructors or content developers. Further, empirically-based conclusions on mobile training are simply not yet possible due to the lack of reported data on learning outcomes (Note only 11 of the 77 sources were empirical mobile studies). For example, a critical analysis of mobile learning projects by Frohberg et al. (2009) identified 102 mobile learning projects, but nearly all were geared to novice learners, such as young pupils, and there were none dealing with learners with extensive previous knowledge. Frohberg et al. (2009) also indicated that only 18 of the 102 projects addressed higher levels of learning beyond the Know (i.e., ability to recall information) and Comprehend (i.e., ability to interpret and summarize information) categories of Bloom's taxonomy (1953).

Benefits and challenges of virtual classrooms. Evidence from the literature review showed that blended strategies incorporating distance learning can provide results (e.g., trainee pass rate, trainee performance) that are equal (Tucker & Goodwin, 2010) or superior to traditional in-class delivery (Coll, Rochera, Mayordomo, & Naranjo, 2007; Lancaster & McQueeney, 2011; Pereira, Pleguezuelos, Molina-Ros, Molina-Tomás, & Masdeu, 2007). The literature review also highlighted specific strategies for enhancing learning outcomes through incorporating distance learning/virtual classroom strategies in a "blended" framework with mobile devices, such as sending real-time assignments via mobile device for students enrolled in a virtual classroom to complete upon delivery (Chen, Chang, & Wang, 2008). The resulting outcomes included increased training completion rates and improved overall training performance.

Benefits and challenges of videogames and virtual worlds. The use of videogames for training is not as novel as the use of virtual worlds (e.g. Second Life). However, many of the purported benefits and challenges associated with these two platforms are similar. Our research noted benefits for both platforms in one of two areas: training motivation and training efficiency. Engagement in a game-based training environment is expected to increase trainee motivation (Mautone et al., 2010) as the result of fully engaging learners and affording them control and choice in a challenging environment (Topolski, Leibrecht, Cooley, Rossi, Lampton, & Knerr, 2010). Similarly, although virtual worlds do not necessarily possess a "game" element, a virtual world environment can potentially increase motivation through the interactive nature of the environment (Chang, Gütl, Kopeinik, & Williams, 2009). Further, both platforms, as a result of their architecture, easily supply learners with timely feedback that has customizable levels of specificity (Mautone et al., 2010).

Videogames and virtual worlds have the ability to create efficient training environments. The ubiquitous nature of these platforms allows learners to train when they want and for as long as they want, which, as noted in the interviews and exemplary elements, gives learners more opportunities to practice and engage in the learning environment, which leads to increased learning (Mautone et al., 2010). Videogames and virtual worlds also offer the opportunity to embed the instructional strategy of narrative storytelling (a method of embedding content in a context-based format to improve recall) into the training event (Hays, Silva, & Richmond, 2011). In addition, as shown in an exemplary element, the use of virtual environments allows for the recreation of potentially dangerous or "high-risk" real-world scenarios, which is an enormous benefit if one considers the cost and risks associated with medical, emergency, and military training (i.e., maneuvering a tank; Tichon, 2007).

Challenges associated with the use of videogames and virtual worlds for training are development costs, linking game play to learning theories and ISD practices, assessment, and evaluation (de Freitas, Rebolledo-Mendez, Liarokapis, Magoulas, & Poulovassilis, 2010). Our interview insights noted that creating a videogame or virtual world increases costs associated with development and thus reduces the return on investment. While leveraging off-the-shelf products, such as popular first-person shooter videogames or Second Life, could greatly reduce costs, they may not align with the learning objectives. The developers of videogames and virtual worlds are not necessarily privy to, or concerned with, the science of learning; their goal is solely the entertainment of the user. Yet, the interview insights note that entertainment alone does little to affect learning if learning objectives, and relevant assessments, are not integrated into the environment. Affective responses merely signify the reception of the environment by the user; the onus of ensuring that learning is occurring is still on the training developers (Okuda et al., 2011).

To ensure learning is occurring, trainers are afforded numerous assessment possibilities in a videogame and virtual world environment; however, the interview insights point out that without careful planning trainers are presented with an overload of learner data. Thus, it is important for training developers to determine the behaviors of interest, the trainee actions indicative of these behaviors in the videogame and virtual world, and ensure that the ability to capture this information exists (Bolstad, Endsley, Costello, & Howell, 2010; Montijo et al., 2010).

Importance of understanding the capabilities of the technology. The interview insights noted many ways that assessments can take advantage of the computational power of training technologies, such as recording structured behaviors while interacting with learning content, conducting after action reviews within virtual environments, conducting assessments on a frequent basis through mobile devices, mining data based on any user activity in virtual worlds, embedding test feedback as part of a training activity, or applying advances in computational linguistics for scoring essays and short responses. Understanding the congruence between the capabilities of the proposed platform (e.g., virtual world), the training objectives (e.g., improve group communication skills), and assessment strategies is vital to the success of a training. Certain capabilities may be available in some platforms and not others. For example, in videogame simulations, duration, quality and/or quantity of certain gameplay actions may be used as indicators of different competencies. A failure to consider the capabilities of the technology could lead to either an inability to assess learners adequately or the missed opportunity to assess in more innovative or complex ways beyond multiple-choice items.

The literature review yielded examples of applications that extend assessment capabilities. These included challenging simulations and measurement tools in medical education and diagnostic testing techniques that measured different levels of acquiring knowledge (Kalyuga, 2006b; Rosen, Salas, Silvestri, Wu, & Lazzara, 2008). These findings would encourage those involved with planning to look beyond the initial layer of what a technology can offer and extend that to assess knowledge and skills at a deeper cognitive level, or at a more frequent schedule. Training developers should also consider using those assessment capabilities for multiple purposes, such as for measuring skill retention.

Provisional Guidelines and Proposed Practices

Presented in Table 7 are the provisional guidelines that were drawn from all of the themes identified in the previous sections. In addition, specific, proposed practices were developed that provide an illustration of how some of the provisional guidelines might be implemented. Because some of the practices are not yet widely implemented and tested, they cannot be declared a best practice. At this point, they can be considered the foundation of a potential best practice.

Table 7

Provisional Guidelines and Proposed Practices

Topic Area	Provisional Guidelines	Proposed Practices
Training Platform Integration	*Guideline 1: To ensure the effectiveness and efficiency of the training, apply principles of the learning sciences to plan the integration of learning experiences across training platforms.*	Practice 1: Involve content, training and technology-platform SMEs throughout the training development process to plan how learning will occur within and across platforms. Practice 2: Evaluate the ability of training content to be included in multiple platforms or transferred from one platform to another to increase the availability of training for learners. Practice 3: Create situated learning opportunities within the platform that are adaptable to the environmental context, or range of contexts, in which the skill will be applied.
Topic Area	**Provisional Guidelines**	**Proposed Practices**
Role of Assessments	*Guideline 2: Develop an assessment strategy and incorporate that strategy into the training framework to ensure assessment(s) provide learner performance data that support the overall goals of the training.* *Guideline 3: Use the same platform for assessment that was used for training if the platform can adequately capture the necessary assessment data to maximize training efficiency.*	Practice 4: Use an assessment map in the development of an assessment strategy to ensure the assessments measure the learning objectives taught in the training and provide useful data for training evaluation.
	Guideline 4: Use assessments to adapt training to learners' proficiency levels and/or style preferences to support a learner-centric environment. *Guideline 5: Employ frequent testing as a means to deliver content, reinforce what was learned, and support adaptive instruction.*	Practice 5: Evaluate the requirements, costs, and usefulness of adaptive tests before including them in an assessment strategy.
	Guideline 6: Implement a holistic method for aggregating and analyzing all sources of social exchange data to assess critical learning objectives within peer- and collaborative-learning scenarios.	Practice 6: Use platforms that have the ability to collect and consolidate social exchange data when that type of data aligns with the learning objectives.
Training Technologies	*Guideline 7: Capitalize on the capabilities of the training technology to assess competencies in alternative ways that are not feasible through traditional platforms.*	Practice 7: Apply knowledge of testing capabilities in emerging platforms when developing the assessment map to determine the most appropriate platforms for assessments.

Discussion

With a specific and narrow focus, this qualitative research sought to uncover exemplars that could inform best practices and guidelines for the integration of three specific learning technology platforms – mobile devices, virtual worlds and videogame simulations, and virtual classrooms – and the use of assessments within these specific platforms. While a very small number of exemplars were found in the literature that met the pre-established criteria, themes were identified from those exemplars, along with the exemplary elements and interview insights that were gathered, that informed the development of provisional guidelines and proposed practices based on the current state of research and practice.

The provisional guidelines that were developed provide an initial framework to assist the Army in the implementation of its Army Learning Model (ALM), and should be viewed as supplementary to other well-established guidelines that are relevant to distance learning and learner-centric approaches. Namely, a subset of well-established and empirically-based guidelines were identified, known as *What Works in Distance Learning* (O'Neil, 2004), that remain applicable to developing and assessing learning content in adult learning centers. These guidelines are presented in Table 8. In addition, the American Psychological Association has published a set of learner-centered psychological principles (APA, 1993; 1997). While these principles address the learner-centric goals of the ALM, they do not address the use or integration of assessments or technology.

Table 8
What Works in Distance Learning Guidelines

Guideline	Description
Multimedia	
Strategies Based on Coherence Principle	Coherence effect: People learn better from multimedia messages when extraneous words, pictures, and sounds are excluded rather than included.
Strategies Based on Modality Principle	Modality effect: People learn better from animation and narration than from animation and on-screen text.
Strategies Based on Multimedia Principle	Multimedia effect: People learn better from corresponding words and graphics (e.g., animation, video, illustrations, and pictures) than from words alone.
Strategies Based on Personalization Principle	Personalization effect: People learn better from multimedia lessons when the words are in conversational style rather than formal style.
Strategies Based on Redundancy Principle	Redundancy effect: People learn better from animation and narration than from animation, narration, and on-screen text.

Table 8 ***(continued)***

Instructional Strategies	
Strategies Based on Providing Worked Examples and Practice	When instruction provides clear (to the learner) and complete procedural "how to" examples of the decisions and actions needed to solve problems and perform necessary tasks to be learned, then learning and transfer will be increased.
Strategies Based on Effective Feedback During Learning	Effective feedback about learning progress results in better learning and transfer of learning to the work environment.
Strategies Based on Increasing Student Motivation: Encouraging Active Engagement and Persistence	Designers can help students to become actively engaged in a course or lesson and to persist or stay "on track" when distracted by helping students connect their personal goals and interests to course goals, by clearly communicating the utility of the course goals (and the risk of not achieving them), and by helping students maintain their confidence in achieving the course goals (by pointing out past successes with similar goals.
Strategies Based on Teaching Causal Principles	When teaching causal principles, learning and transfer to the job will be more effective the more that the instructional presentation provides a statement about the cause and resulting effects, provides instruction using a worked, prototypical example drawn from the application setting, and helps the learner to first elaborate the elements and sequence of the causal chain and then to apply it to gradually more novel and complex examples.
Assessment	
Cognitive Demands Strategies	Assessment specifications should explicitly reference both the models of cognitive demand in the task, (e.g., knowledge understanding or problem solving) and the cognitive requirements of desired performance in the specific content area.
Domain Representation Strategies	Tests must contain adequate sampling of items or tasks that are representative of the content domain to be assessed.
Formative Assessment Strategies	Tests given during instruction should provide information for feedback and motivation to the learner, guide the program to provide needed help, and give the instructional designer information about program strengths and weaknesses.

Note. These strategies can be found in O'Neil (2004).

Research Questions Revisited

Given the outcomes of the research reported herein, it is useful to revisit ARI's overarching research questions to see how these outcomes and the qualitative data, in general, address the questions.

How should assessments be designed, delivered, and otherwise used to maximize Soldier training? Assessments should be carefully planned to ensure that they provide the necessary information at the right time to support the tailored instruction that is required by the ALM. Given the increased amount of data that can be captured and analyzed with the platforms of interest, it is critical to determine beforehand which data will be useful in assessing the learning objectives, how often that data should be captured, and how that data should be analyzed. An assessment map is a recommended approach for developing an assessment strategy.

Assessments should be delivered through the same platform(s) that is used for training to maximize efficiency, minimize learner disruption, and test in the same manner in which the material is taught. They should be used to not only to assess that learning has occurred to standard, but also to adapt training to a learner's preferences and proficiency level. In addition, assessments can and should be viewed as a mechanism for delivering training content to increase learning opportunities.

The capabilities within learning technologies should be used to assess learning objectives in more complex and innovative ways than may be possible in traditional training platforms. Assessment data that provides information on duration, quality, or quantity of behaviors can be used to assess learning and can be done so in a manner that is hidden to the learner, making learning and assessment seamless. In addition, the platforms of interest also provide the capability to capture and analyze social exchanges that can provide assessment of competencies such as teamwork and collaboration.

How should adaptive assessments be conducted? Adaptive tests can provide fine-tuned tailoring of training, but there are considerations for their use. Recommendations regarding the use of adaptive tests are to evaluate their usefulness for the particular training effort and to weigh that against the requirements of adaptive tests. If the training content or the training platform does not allow for a high degree of tailoring to the learner, adaptive testing may be an unnecessary expense.

How often should assessments be conducted? Frequent assessment should be included in the assessment strategy to support tailored instruction. Pre- and post-tests within each platform should be included to provide evidence of learning in each platform and to tailor training that the learner receives in the next platform. To provide more fine-tuned tailoring of training within a platform, progress tests need to be included in the assessment strategy. Our research supports the use of progress tests because not only does frequent assessment allow for training to be tailored to the learner's proficiency level, it also can serve as a means to deliver content and reinforce training material.

What are Soldiers' preferences for training on technology-based platforms? While our research did not specifically look for information on Soldier's preferences, the findings suggest that the platforms of interest should be motivating to Soldiers. These platforms offer the potential for increased engagement and collaboration with others, situated learning opportunities that are adaptable to the range of operational environments, timely feedback, and flexibility to train at the learner's convenience.

How effective is training that is delivered through technology-based platforms? Evidence from the exemplars indicates the platforms of interest can be effective in training. The first two exemplars demonstrated gains in learning from pre- and post-tests for training delivered through the platforms of interest. In addition, the first exemplar demonstrated gains over traditional classroom training. The third exemplar demonstrated higher performance when training assessed learner progress throughout and provided reminders to learners based on their progress through technology platforms.

A critical factor in ensuring the effectiveness of training delivered through any platform is the use of principles of the learning sciences in the planning and development of the training. The learning objectives need to be the driving factor in the determination of which platforms to use and which content should be included in each platform. To assist with these decisions, the use of decision tools and SMEs are recommended. It was also recommended, based on our research, that technology platforms be used in a way to create situated learning opportunities that are similar to the environmental context, or adaptable to the range of contexts, in which the skill will be applied. Selecting and using the appropriate technology to increase the fidelity of the training to the actual environmental context can increase skill transfer and retention.

What are best practices for delivering and developing training evaluations to maximize the benefits of leveraging these emerging technologies? Training evaluations are most effective when planned in conjunction with the training assessment strategy following a training needs analysis. The rationale for this process is that: 1) assessment and evaluation should be informed by the training needs of an organization and 2) training evaluators can leverage data that is used to assess individual and/or team performance to evaluate the effectiveness of training, a practice highlighted by Exemplar 1 (Montijo et al., 2010). Therefore, it is important to consider training evaluation during assessment planning and development. Further, careful thought should be given to the criterion used to determine overall training effectiveness, which then should be aligned with the assessment strategies.

When planning for training evaluation, developers need to consider the capabilities of the learning technology platform being used and how those capabilities may or may not impact the effectiveness of the training. For example, if training content is sound, but delivered through the wrong learning technology platform (e.g., providing training content that necessitates visual activities that are not suited for a small screen on a mobile phone), the training evaluation data may indicate that learning did not occur. However, that general evaluation of the training (i.e., Did training work?) would be confounded by the technology platform. An evaluation of trainee feedback may allow the developers to discern the issue, but the effectiveness of the content would still be unknown. Developers should consider this issue when planning for the training

evaluation and seek to identify ways to assess both the effectiveness of the content and the platform.

Limitations

The research reported herein had several limitations. First, the definition of an exemplar was very narrow given the specific vision of the SCALE prototype training. The research was focused on the operational use of assessments within the three specific platforms of interest, which greatly restricted the sample of potential exemplars. Much of the published literature on incorporating assessment in the platforms of interest reports on prototype or "proof-of-concept" trainings, generally in educational settings (i.e., Kindergarten to 12th grade).

Second, we sought empirical evidence of the effectiveness of reported trainings or techniques in order to support the identification of best practices and guidelines. Many of the sources either provided no empirical evidence or provided weak empirical support. Of the 77 sources in the database, only 18 included a comparison group. The highest score that any source achieved on the Relevance and Empirical Strength scale was 19 out of 49 points, with an average of nine points (SD = 4.95). This finding indicates the need for more rigorous empirical research in this area.

In addition, the lack of empirical evidence in the 77 sources limited the ability to conduct any type of quantitative analysis to support the identification of best practices or guidelines. Only 31 sources had a design and included a variable where effect size could have been calculated. Of those 31 sources, only 25 included actual effects sizes or enough detail to compute an effect size.

Future Research

Before the provisional guidelines can be fully established, there needs to be a stronger foundation of evidence and support by both researchers and the practitioners responsible for their implementation. Future research should include rigorous empirical studies of blended training with an operational focus, that uses multiple platforms, and assesses learners prior to training, during training, immediately after training, and then in pre-determined time periods following training to assess retention. These assessments might be used in an adaptive manner to determine the best training content to initially deliver to learners as well as what training content to deliver as learners progress. The choice of training platform, assessment method, and adaptive technique should be linked to the desired competencies and behaviors that represent those competencies. This is one of many studies that could potentially test the usefulness of the provisional guidelines. Until then, the provisional guidelines, along with the established guidelines, form a foundation for a reasoned approach to using and integrating assessments and emerging technologies in future Army training.

References

*Allan, K. (1993). Computerized course teaches interviewing. *Personnel Journal, 72*(6), 66.

APA Task Force on Psychology in Education (1993). *Learner-centered psychological principles: Guidelines for school redesign and reform*. Washington, DC: American Psychological Association and Mid- Continent Regional Educational Laboratory.

APA Work Group of the Board of Educational Affairs (1997). *Learner-centered psychological principles: A framework for school reform and redesign*. Washington, DC: American Psychological Association.

*Behrend, T. S. (2011). *Recruitment and selection in the virtual world* [Presentation Slides]. Presentation to the Personnel Testing Council February, 2011 Luncheon. Arlington, VA.

*Behrens, J. T., Mislevy, R. J., DiCerbo, K. E., & Levy, R. (2010). *An evidence centered design for learning and assessment in the digital world*. (CRESST Report 778). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Bloom, B. S. (1953). Thought processes in lectures and discussions. *Journal of General Education 7,* 160-169.

*Bolstad, C. A., Endsley, M. R., Costello, A. M., & Howell, C. D. (2010). Evaluation of computer-based situation awareness training for general aviation pilots. *The International Journal of Aviation Psychology, 20*, 269-294.

*Bowling, S. R., Khasawneh, M. T., Kaewkuekool, S., Jiang, X., & Gramopadhye, A. K. (2008). Evaluating the effects of virtual training in an aircraft maintenance task. *International Journal of Aviation Psychology, 18*, 104-116.

*Byun, S., & Mills, J. E. (2011). Exploring the creation of learner-centered e-training environments among retail workers: A model development perspective. *CyberPsychology, Behavior & Social Networking, 14*, 65-69.

*Cameron, B., & Dwyer, F. (2005). The effect of online gaming, cognition and feedback type in facilitating delayed achievement of different learning objectives. *Journal of Interactive Learning Research, 16,* 243-258.

*Chang, V., Gutl, C., Kopeinik, S., & Williams, R. (2009). Evaluation of collaborative learning settings in 3D virtual worlds. *iJet, 4*, 6-17.

*Chen, C.-M., & Hsu, S.-H. (2008). Personalized intelligent mobile learning system for supporting effective English learning. *Educational Technology & Society, 11*, 153-180.

Note: References marked with an asterisk were included in the literature review.

*Chen, G. D., Chang, C. K., & Wang, C. Y. (2008). Ubiquitous learning website: Scaffold learners by mobile devices with information-aware techniques. *Computers & Education, 50*, 77-90.

Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validity. *Personnel Psychology, 63*, 83-117.

*Clough, G., Jones, A. C., McAndrew, P., & Scanlon, E. (2008). Informal learning with PDAs and smartphones. *Journal of Computer Assisted Learning, 24*, 359-371.

*Coll, C., Rochera, M. J., Mayordomo, R. M., & Naranjo, M. (2007). Continuous assessment and support for learning: An experience in educational innovation with ICT support in higher education. *Electronic Journal of Research in Educational Psychology, 5*, 783-804.

*de Freitas, S., Rebolledo-Mendez, G., Liarokapis, F., Magoulas, G., & Poulovassilis, A. (2010). Learning as immersive experiences: Using the four-dimensional framework for designing and evaluating immersive learning experiences in a virtual world. *British Journal of Educational Technology, 41*, 69-85.

*Dean, C., Webb, S., Keeney, M., Day, E., Terry, R., & Alicia, T. (2011). Item response theory adapts training to disparately skilled trainees. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference*, (pp. 3009-3018). Orlando, FL.

*Delacruz, G. C., & Iseli, M.R. (2011). Assessment architectures to support development and validation of adaptive training. [Presentation slides]. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference*, (pp. 1089-1154). Orlando, FL.

*Delfino, M., & Persico, D. (2007). Online or face-to-face? Experimenting with different techniques in teacher training. *Journal of Computer Assisted Learning, 23*, 351-365.

*Desai, M., Richards, T., & Eddy, J. P. (2000). A field experiment: Instructor-based training vs. computer-based training. *Journal of Instructional Psychology, 27*, 239-243.

*Feldman, M. J., Barnett, G. O., Link, D. A., Coleman, M. A., Lowe, J. A., & O'Rourke, E. J. (2006). Evaluation of the clinical assessment project: A computer-based multimedia tool to assess problem-solving ability in medical students. *Pediatrics, 118*, 1380-1387. doi: 10.542/peds.2006-0326

*Frohberg, D., Goth, C., & Schwabe, G. (2009). Mobile learning projects – a critical analysis of the state of the art. *Journal of Computer Assisted Learning, 25*, 307-331.

*Grant, S., & Clerehan, R. (2011). Finding the discipline: Assessing student activity in *Second Life*. *Australasian Journal of Educational Technology, 27*(Special Issue, 5), 813-828.

Note: References marked with an asterisk were included in the literature review.

*Gu, X., Gu, F. & Laffey, J. M. (2011). Designing a mobile system for lifelong learning on the move. *Journal of Computer Assisted Learning, 27*, 204-215.

*Guzman, E., Conejo, R., & Perez-de-la-Cruz, J. (2007). Improving student performance using self-assessment tests. *Intelligent Educational Systems, 22*, 46-52.

*Haag, J. (2011, Dec). From eLearning to mLearning: The effectiveness of Mobile Course Delivery. *Proceedings from the Interservice/Industry, Training, Simulation, and Education Conference*, Orlando, FL.

*Hassan, I. S., Ismail, M. A., & Mustapha, R. (2010). The effect of integrating mobile and CAD technology in teaching design process for Malaysian polytechnic architecture student in producing creative product. *TOJET: The Turkish Online Journal of Educational Technology, 9*, 162-172.

*Haynes, J. A. (2010). Instructional system design & simulation-based instruction. [Presentation slides]. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference*, (pp. 1449-1510). Orlando, FL.

*Hays, M. J., Silva, T. M., & Richmond, T. (2011). Assessing learning from a mixed-media, mobile counter IED trainer. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference,* (pp. 2281-2289). Orlando, FL.

Holden, C. L., & Sykes, J. M. (2011). Leveraging mobile games for place-based language learning. *International Journal of Game-Based Learning, 1*, 1-18.

*Hwang, G-J., & Chang, H-F. (2011). A formative assessment-based mobile learning approach to improve the learning attitudes and achievements of students. *Computers & Education, 56*, 1023-1031.

*Iseli, M. R., Koenig, A. D., Lee, J. J., & Wainess, R. (2010). *Automated assessment of complex task performance in games and simulations*. Retrieved April 25, 2012, from The University of California Los Angeles, National Center for Research on Evaluation, Standards, & Student Testing Web site: http://www.cse.ucla.edu/products/reports.php?action=search&query=775.

*Kalyuga, S. (2006a). Assessment of learners' organized knowledge structures in adaptive learning environments. *Applied Cognitive Psychology, 20*, 333-342.

*Kalyuga, S. (2006b). Rapid assessment of learners' proficiency: A cognitive load approach. *Educational Psychology, 26*, 735-749.

*Kearns, S. (2011). Online single-pilot resource management: Assessing the feasibility of computer-based safety training. *International Journal of Aviation Psychology, 21*, 175-190.

Note: References marked with an asterisk were included in the literature review.

*Klein, H. J., Noe, R. A., & Chongwei, W. (2006). Motivation to learn and course outcomes: The impact of delivery mode, learning goal orientation, and perceived barriers and enablers. *Personnel Psychology, 59,* 665-702.

Knowles, M. (1980). *The modern practice of adult education: From pedagogy to andragogy*: Wilton, CT: Association Press.

*Krätzig, G. P., Bell, G., Groff, R., & Ford, C. (2010). Simulator emergency vehicle operation: Efficiencies and skill transfer. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference,* (pp. 3042-3052). Orlando, FL.

*Lai, C. H., Yang, J. C., Chen, F. C., Ho, C. W., & Chan, T. W. (2007). Affordances of mobile technologies for experiential learning: The interplay of technology and pedagogical practices. *Journal of Computer Assisted Learning, 23,* 326-337

*Lancaster, J. W., & McQueeney, M. L. (2011). From the podium to the PC: A study on various modalities of lecture delivery within an undergraduate basic pharmacology course. *Research in Science & Technological Education, 29,* 227-237.

Lave, J. (1988). *Cognition in Practice: Mind, mathematics and culture in everyday life*. Cambridge, UK: Cambridge University Press.

*Lewis, N. J., & Orton, P. Z. (2006). Blending learning for business impact: IBM's case for learning success. In C. J. Bonk & C. R. Graham (Eds.), *The handbook of blended learning: Global perspectives, local designs* (pp. 61-75). San Francisco: Pfeiffer.

*Liu, T. Y. (2009). A context-aware ubiquitous learning environment for language listening and speaking. *Journal of Computer Assisted Learning, 25,* 515-527.

*Looi, C. K., Zhang, B., Chen, W., Seow, P., Chia, G., Norris, C., & Soloway, E. (2011). 1:1 mobile inquiry learning experience for primary science students: A study of learning effectiveness. *Journal of Computer Assisted Learning, 27,* 269-287.

*Lu, M. (2008). Effectiveness of vocabulary learning via mobile phone. *Journal of Computer Assisted Learning, 24,* 515-525.

*Mautone, T., Spiker, A., Karp, M. R., & Conkey, C. (2010). Using games to accelerate aircrew cognitive training. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference,* (pp. 1898-1909). Orlando, FL.

*McConatha, D., & Praul, M. (2007, August). *Mobile learning in the classroom: An empirical assessment of a new tool for students and teachers.* Paper presented at the meeting of the Society for Applied Learning Technology's, Arlington, VA.

Note: References marked with an asterisk were included in the literature review.

*McDowell, P., Johnson, R. E., Freeman, J., Roberts, M., & Horn, Z. (2011). Building a game to educate senior officers in counter-piracy. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference,* (pp. 2301-2311). Orlando, FL.

*Mebane, M., Porcelli, R., Iannone, A., Attanasio, C., & Francescato, D. (2008). Evaluation of the efficacy of affective education online training in promoting academic and professional learning and social capital. *International Journal of Human-Computer Interaction, 24*, 68-86.

*Metzler-Baddeley, C. R. J. (2009). Does adaptive training work? *Applied Cognitive Psychology, 23*, 254-266.

*Mitchell, S., Heyden, R., Heyden, N., Schroy, P., Andrew, S., Sadikova, E., & Wiecha, J. (2011). A pilot study of motivational interviewing training in a virtual world. *Journal of Medical Internet Research, 13*, e77.

*Montijo, G. A., Spiker, V. A., & Nullmeyer, R. (2010). Training interventions to reduce predator crew errors. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference,* (pp. 3021-3031). Orlando, FL.

*Nash, P., & Shaffer, D. W. (2011). Mentor modeling: The internalization of modeled professional thinking in epistemic game. *Journal of Computer Assisted Learning, 27,* 173-189.

Norris, C., & Soloway, E. (2004). Envisioning the handheld centric classroom. *Journal of Educational Computing Research, 30*, 281-294.

*Ochoa, J. D. K. (2012). Using a virtual training program to train community neurologist on EEG reading skills. *Teaching & Learning in Medicine, 24*, 26-28.

*Okuda, H., Arcaro, L., & Gaught, B. (2011). Understanding the healthcare simulation development lifecycle [Presentation slides]. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference, (*pp. 483-519). Orlando, FL.

O'Neil, H. (2004). *What works in distance learning: Guidelines.* Greenwich, Connecticut: Information Age Publishers.

*O'Neil, H.F., Baker, E. L., Wainess, R., Chen, C., Mislevy, R., & Kyllonen P.(2004). *Final report on plan for the assessment and evaluation of individual and team proficiencies developed by the DARWARS environments*. Sherman Oakes, CA: Advance Design Information. (DTIC No. ADA432802).

*Owne, H., Mugford, B., Follows, V., & Plummer, J. L. (2006). Comparison of three simulation-based training methods for management of medical emergencies. *Resuscitation, 71*, 204-211.

Note: References marked with an asterisk were included in the literature review.

*Pereira, J. A., Pleguezuelos, E., Merí, A., Molina-Ros, A., Molina-Tomás, M. C., & Masdeu, C. (2007). Effectiveness of using blended learning strategies for teaching and learning human anatomy. *Medical Education, 41*, 189-195.

Pleban, R. J., Eakin, D. E., Salter, M. S., & Matthews, M. D. (2001). Training and assessment of decision-making skills in virtual environments* (ARI Research Report 1767). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

*Pokorny, R., Haynes, J., & Gott, S. (2010). Practical assessment in complex environments. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference,* (pp. 1786-1795). Orlando, FL.

Roschelle, J., & Pea, R. (2002) A walk on the WILD side: How wireless handhelds may change computer- supported collaborative learning. *International Journal of Cognition and Technology,1*, 145-168.

*Rosen, M. A., Salas, E., Silvestri, S., Wu, T. S., & Lazzara, E. H. (2008). A measurement tool for simulation-based training in emergency medicine: The simulation module for assessment of resident targeted event responses (SMARTER) Approach. *Society for Simulation in Healthcare, 3*, 170-179.

*Rosen, M. A., Salas, E., Weaver, S. J., Lazzara, E. H., King, H. B., & Robinson, D. (2010). [Presentation slides]. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference,* (pp. 974-1060). Orlando, FL.

*Ross, W. A., & Kobus, D. A. (2011). Case-based next generation cognitive training solutions. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference,* (pp. 2858-2864). Orlando, FL.

*Salmon, G., Ming, N., & Palitha, E. (2010). Developing a five-stage model of learning in second life. *Educational Research, 52*, 169-182.

*Sancho, P., Moreno-Ger, P., Fuentes-Fernandez-Manjon, R., & Fernandez-Manjon, B. (2009). Adaptive role playing games: An immersive approach for problem based learning. *Educational Technology & Society, 12*, 110-124.

*Schoppek, W. M. (2010). Enhancing arithmetic and word-problem solving skills efficiently by individualized computer-assisted practice. *Journal of Educational Research, 103*, 239-252.

*Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology, 64*, 489-528.

*Sitzmann, T., Kraiger, K., Stewart, D., & Wisher, R. (2006). The comparative effectiveness of web-based and classroom instruction: A meta-analysis. *Personnel Psychology, 59*, 623-664.

Note: References marked with an asterisk were included in the literature review.

*Smikowski, J., Dewane, S., Johnson, M. E., Brems, C., Bruss, C., & Roberts, L. W. (2009). Community-based participatory research for improved mental health ethics & behavior. *Ethics and Behavior, 19,* 461-478.

*So, H. J. (2009). When groups decide to use asynchronous online discussions: Collaborative learning and social presence under a voluntary participation structure. *Journal of Computer Assisted Learning, 25,* 143-160.

Soloway, E., Norris, C., Blumenfeld, P., Fishman, B., Krajcik, J., & Marx, R. (2001). Log on education: Handheld devices are ready-at-hand. *Communications of the ACM, 44,* 15-20.

*Sotomayor, T. (2008). Evaluating tactical combat casualty care training treatments effects on combat medic trainees in light of select human descriptive characteristics. (Doctoral dissertation, George Washington University, 2008). *Dissertation Abstracts International, 70(11-A),* 4257.

*Swezey, R. W., Hutcheson, T. D., & Swezey, L. L. (2000). Development of a second-generation computer-based team performance assessment technology. *International Journal of Cognitive Ergonomics, 4,* 163-170.

*Tichon, J. (2007). Training cognitive skills in virtual reality: Measuring performance. *CyberPsychology & Behavior, 10,* 286-289.

*Tobias S., & Fletcher J. D. (2007). What research has to say about designing computer games for learning. *Educational Technology*, *47*, pp. 20–29.

*Topolski, R., Leibrecht, B., Cooley, S., Rossi, N., Lampton, D., & Knerr, B. (2010). *Impact of game-based training on classroom learning outcomes*, (Technical Report 2010-01). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences (DTIC No. ADA531677).

TRADOC (2011). *The U. S. Army Learning Concept for 2015. TRADOC Pamphlet 525-8-2.* Retrieved from http://www.tradoc.army.mil/tpubs/pams/tp525-8-2.pdf.

*Traxler, J., & Kukulska-Hulme, A. (2005). *Evaluating Mobile Learning: Reflections on Current* Practice Paper presented at the mLearn 2005 4th World Conference on mLearning. Mobile Technology: The future of learning in your hands. Cape Town, South Africa.

Triantafillou, E., Georgiadou E., & Economides A. A. (2008). The design and evaluation of a computerized adaptive test on mobile devices. *Computers & Education, 50,* 1319-1330.

*Tsai, C.-W. (2011). Achieving effective learning effects in the blended course: A combined approach of online self-regulated learning and collaborative learning with initiation. *CyberPsychology, Behavior & Social Networking, 14,* 505-510.

Note: References marked with an asterisk were included in the literature review.

Tucker, J. S. (2010). *Mobile learning approaches for U.S. Army training* (Research Note 2010-07). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

*Tucker, J. S., & Goodwin, G. A. (2010). Soldier performance following distributed and traditional digital skills training. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference,* (pp. 2958-2968). Orlando, FL.

U. S. Department of Education. (1987). *What works: Research about teaching and learning,* Washington, D.C.: Author

*Vavoula, G. N., & Sharples, M. (2008). *Challenges in evaluating mobile learning*. Paper presented at the annual meeting of the mLearn Conference, Wolvenhampton, United Kingdom.

*Wainess, R., Koenig, A., & Kerr, D. (2010). Aligning instruction and assessment with game and simulation design. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference,* (pp. 1796-1805). Orlando, FL

*Walker, B. L., & Harrington, S. S. (2003). Computer-based and instructor-led injury prevention training for board and care staff. *Educational Gerontology, 29*, 475-491.

*Walker, B. L., & Harrington, S. S. (2008). Computer-based instruction and the web for delivering injury prevention training. *Educational Gerontology, 34,* 691-708.

*Wang, A. Y., Newlin, M. H., & Tucker, T. L. (2001). A discourse analysis of online classroom chats: Predictors of cyber-student performance. *Teaching of Psychology, 28*, 222-226.

Note: References marked with an asterisk were included in the literature review.

APPENDIX A
LIST OF DATABASE FIELDS

Database Variables

Field Name	Type	Description
Basic Information Tab		
ID	Long Integer	Unique record ID
Rater Initials	Text	Initials of coder
Short Title	Text	Short title of reference
Exemplar	Yes/No/Maybe	Dose record meet exemplar criteria
Keyword(s)	Text	Reference listed key words
Sponsor	Text	Identifies who the training was developed for
Literature Review Fields Tab		
APA Citation	Text	APA formatted citation of reference
Abstract	Text	Reported abstract
Organization Reporting	Text	Organization of 1st author
Literature Type	Text	What medium was reference published
Date of Publication (if applicable)	Long Integer	Date of publication
On-going Activity	Yes/No	Does author indicate on-going research activities
Training Fields Tab		
Training Audience	Long Integer	Who is target of training effort
Training Platform		What platform was used for the training effort
Mobile	Yes/No	
Videogames	Yes/No	
Virtual World	Yes/No	
Traditional Classroom	Yes/No	
Virtual Classroom	Yes/No	
Other	Yes/No	
Description of Training Sample/Population	Text	Description of the sample participating in training activities
Description of Training Effort	Text	Description of the training effort(s) reported
Description of Training Platform(s)	Text	Description of the training platform
Training/Assessment Developer	Text	Description of who developed training & assessment platform(s)
Training Software		What type of software was used in training
COTS	Yes/No	
Custom	Yes/No	
Unknown/Not Reported	Yes/No	
Training Content		What was the focus of the training

Individual	Yes/No	
Team	Yes/No	
Procedural	Yes/No	
Cognitive	Yes/No	
Psychomotor	Yes/No	
Affective	Yes/No	
Assessment Fields Tab		
Description of Platform-Specific Assessments	Text	Provide description of assessments used for each platform
Description of Cross Platform Assessment Strategy	Text	Provide description of assessment strategies used across multiple platforms
Description of Cross-Platform Assessments	Text	Provide description of assessments used across platforms
Description of Cross Platform Assessment Issues/Challenges	Text	Discuss noted challenges or issues in using assessments across multiple platforms
Category of Measurement		Identify the level of assessment
K1 -Reaction	Yes/No	
K2 -Learning Outcomes (General Use)	Yes/No	
K2 -LO-Multiple-Choice Test	Yes/No	
K2 -LO-Situational Judgment test	Yes/No	
K2 -LO-Adaptive Test	Yes/No	
K3 -Behavior (General Use)	Yes/No	
K3 -Behavior-Individual Performance in a Simulation or Exercise	Yes/No	
K3 -Behavior-Team-based Performance in a Simulation or Exercise	Yes/No	
K4 -Organizational Results	Yes/No	
Training Efficiency	Yes/No	
Training Cost	Yes/No	
Attrition	Yes/No	
Adaptive Instruction	Yes/No	
Knowledge or Skill Retention	Yes/No	
Research Design		What type of research design was used in the record
Quasi-/Experimental	Yes/No	
Matched Sample	Yes/No	
Comparison Group	Yes/No	
Within Subjects	Yes/No	
Between Subjects	Yes/No	
Mixed Subjects	Yes/No	

Other	Text	
Evaluation Fields Tab		
Description of Evaluation Strategy	Text	Describe how evaluation of training was observed (hypotheses testing; formative feedback; etc.)
Evaluation and Validity Shortcomings	Text	Describe shortcomings/limitations of the report
Conclusions/Recommendations	Text	Provide recommendations and conclusions noted by author
Sample Size		Provide specifics about sample used in study
Demographics Reported	Check-box	
Sample Size Comparison	Long Integer	
Sample Size Treatment	Long Integer	
Randomly Assigned Treatment	Yes/No	
Testing Frequency		Identify the frequency of testing within study
Pre-test	Yes/No	
Single Post-test	Yes/No	
Progress-test(s)	Yes/No	
Multiple post-tests	Yes/No	
Reliability of Instruments		Report reliability of instruments used in study
Instrument Name	Text	
Not Reported	Check-box	
Reported Value	Long Integer	
Effect Size of Instruments		Report effect size or whether effect size can be computed
Instrument Name	Text	
Reported Effect Size	Long Integer	
Not Reported	Check-box	
Effect Size Computable	Check-box	

APPENDIX B
INTERVIEW PROTOCOLS

Protocol for Training Developers

Name/Title of Training Effort:

SME Name:

SME Title:

SME Organization/Institution:

Type of Institution (e.g., Military, Higher Education, For-profit private, Non-profit private)

1) Please describe your training effort in terms of:

- Goals
- Target populations (i.e., potential trainees)
- Type of training (individual, collective, combined)
- Training content (introductory, advanced, refresher, etc.)
- Expected/planned outcomes

2) Which training platform formats (i.e., mobile, virtual, videogame-based) were implemented in your effort?

- Why were these selected?
- What evidence supported your decision-making?
- What was the expected utility (i.e., payoff) of the platforms?
- In addition to these platforms, were any additional training platforms utilized? (e.g., Live Classroom)

3) How were training activities developed from a technical perspective?

- Systems and applications (e.g., game engine, mobile device platform) utilized?
- Developed in-house or by external contractor?
- Where is the system housed (e.g., housed or hosted locally, hosted externally by contractor, hosted externally through other company/organization)?
- Did students need to provide their own learning devices (e.g., cell phones, laptops, game box)?

4) How were training activities developed from an ISD perspective? Answer for the following for each training platform incorporated:

- How was the determination made for which training platforms would be used (i.e., media selection)?
- What instructional strategies were developed/incorporated (e.g., direct instruction, indirect instruction, experiential learning, independent study, interactive instruction)?
- What types of learning were targeted (e.g., cognitive, affective, psychomotor, procedural)?
- What was the time-frame and duration of the training (in days)?

- What was the target group for instruction (e.g., individual and collective or small team instruction)?
- Please briefly explain how was training content/activities determined and developed.

5) Please describe the individual learner/participant assessments (i.e., strategies to track individual progress/achievement) that were incorporated into the training effort.

For each training platform module:

- Describe any assessments utilized for platform-specific training (e.g., mobile, virtual, videogame-based):
    - Format (i.e., Paper-and-pencil, computer/mobile device-based, performance-based in live environment, performance based in live simulation).
    - Timing of assessment (e.g., pre/post-test, single point during training, multiple re-assessment during training)?
    - Structure of the assessment items (e.g., multiple-choice, situational judgment, job simulation, interview/de-briefing questions, instructor performance assessment, and team-member performance assessment items).
- Were any validation activities performed? If so, what did they consist of?

Comprehensive/ cross- platform assessment:

- Describe any comprehensive (e.g., summary end-of-course) or cross-platform (e.g., mid-point assessment) training assessments
    - Format (i.e., Paper-and-pencil, computer/mobile device-based, performance-based in live environment, performance based in live simulation).
    - Timing of assessment (e.g., pre/post-test, single point during training, multiple re-assessment during training).
    - Structure of the assessment items (e.g., multiple-choice, situational judgment, job simulation, interview/de-briefing questions, instructor performance assessment, team-member performance assessment items).
- Please describe how individual assessments (e.g., by training platform) are coordinated to assess trainee progress/achievement
    - Does platform-specific assessment information contribute to training placement or tailoring of instruction (e.g. aptitude-treatment interaction) in subsequent modules? If so, how is this executed?

- Were assessment validation activities performed? If so, what did they consist of? What were the results?

- How is assessment information achieved, stored? What is the planned future use (e.g., incorporated into a learning management system)?

    - If a database or LMS is used, please describe the system to include whether a customer-made or commercial tool is used.

6) Training Evaluation (i.e., strategies for determining training effectiveness at the course-level)

- Describe any evaluation strategies that took place to assess at the training course level (e.g., reaction, learning, behavior, results).

- What evaluation results did you have?

- To what extent did the evaluation influence the instructional path, such as with adaptive testing?

- How were evaluation results utilized?

7) Results and Future Plans

- Describe the degree to which the training and assessment strategies used met (or did not meet) expectations and produce planned results.

- What were the 'challenges' that resulted for training and assessment?

    - From a technical development and/or instructional design standpoint

- What innovations (if any) were developed to realize goals?

- What 'lessons learned' (e.g., common pitfalls, low effort/high impact solutions) can you pass along as a recommendation to someone planning to develop and implement a similar training and assessment strategy?

    - What specific suggestions do you have for developing effective cross-platform assessments in a blended learning environment?

- What future plans (e.g., modifications) do you have for your program?

Protocol for Training Program Administrators & Academics

SME Name:

SME Title:

SME Organization/Institution:

Type of Institution (e.g., Military, Higher Education, For-profit private, Non-profit private)

1) Based on your experience how would you go about implementing a cross-platform, individualized training strategy that depends on systematic learner assessments to optimize training effectiveness/efficiencies?

2) Please describe your recent experiences with relevant operational training:

- Goals
- Target populations (i.e., potential trainees)
- Type of training (individual, collective, combined)
- Training content (introductory, advanced, refresher, etc.)
- Expected/planned outcomes

3) Which training platform formats (i.e., mobile, virtual, videogame-based) were implemented in your effort(s)?

- Why were these selected?
- What evidence supported your decision-making?
- What was the expected utility (i.e., payoff) of the platforms?
- In addition to these platforms, were any additional training platforms utilized? (e.g., Live Classroom)

4) How were training activities developed from a technical perspective?

- Systems and applications (e.g., game engine, mobile device platform) utilized?
- Developed in-house or by external contractor?
- Where is the system housed (e.g., housed or hosted locally, hosted externally by contractor, hosted externally through other company/organization)?
- Did students need to provide their own learning devices? (cell phones, laptops, game box)

5) How were training activities developed from an ISD perspective? Answer for the following for each training platform incorporated:

- What instructional strategies were developed/incorporated (e.g., direct instruction, indirect instruction, experiential learning, independent study, interactive instruction)?
- What types of learning were targeted? (e.g., cognitive, affective, psychomotor, procedural)
- What was the time-frame and duration of the training (in days)?

- What was the target group for instruction? (e.g., individual and collective or small team instruction)

6) Please describe the assessments that were incorporated.

For each training platform module:

- Describe any assessments utilized for platform-specific training (e.g., mobile, virtual, videogame-based):
    - Format (i.e., Paper-and-pencil, computer/mobile device-based, performance-based in live environment, performance based in live simulation).
    - Timing of assessment (e.g., pre/post-test, single point during training, multiple re-assessment during training)?
    - Structure of the assessment items (e.g., multiple-choice, situational judgment, job simulation, interview/de-briefing questions, instructor performance assessment, team-member performance assessment items).
- Were any validation activities performed? If so, what did they consist of?

Comprehensive/ cross- platform assessment:

- Describe any comprehensive (e.g., summary end-of-course) or cross-platform (e.g., mid-point assessment) training assessments
    - Format (i.e., Paper-and-pencil, computer/mobile device-based, performance-based in live environment, performance based in live simulation).
    - Timing of assessment (e.g., pre/post-test, single point during training, multiple re-assessment during training).
    - Structure of the assessment items (e.g., multiple-choice, situational judgment, job simulation, interview/de-briefing questions, instructor performance assessment, team-member performance assessment items).
- Please describe how individual assessments (e.g., by training platform) are coordinated to assess trainee progress/achievement.
    - Does platform-specific assessment information contribute to training placement or tailoring of instruction (e.g. aptitude-treatment interaction) in subsequent modules? If so, how is this executed?
- Were assessment validation activities performed? If so, what did they consist of? What were the results?

- How is assessment information achieved, stored? What is the planned future use (e.g., incorporated into a learning management system)?
    - If a database or LMS is used, please describe the system to include whether a customer-made or commercial tool is used.

7) Training Evaluation

- Describe any evaluation strategies (e.g., reaction, learning, behavior, results) used.

- What evaluation results did you have?

- To what extent did the evaluation influence the instructional path, such as with adaptive testing?

- How were evaluation results utilized?

8) Results and Future Plans

- Describe the degree to which the training and assessment strategies used met (or did not meet) expectations and produce planned results.

- What were the 'challenges' that resulted for training and assessment?
    - From a technical development and/or instructional design standpoint

- What innovations (if any) were developed to realize goals?

- What 'lessons learned' (e.g., common pitfalls, low effort/high impact solutions) can you pass along as a recommendation to someone planning to develop and implement a similar training and assessment strategy?
    - What specific suggestions do you have for developing effective cross-platform assessments in a blended learning environment?

- What upcoming plans do you have to complete training and assessment projects (or modify existing projects) in the next 18 months?

APPENDIX C
LIST OF LITERATURE REVIEW EXEMPLARS AND EXEMPLARY ELEMENTS

EXEMPLAR 1: Training Interventions to Reduce Air Force Predator Crew Errors

Citation:

Montijo, G. A., Spiker, V. A., & Nullmeyer, R. (2010). Training interventions to reduce predator crew errors. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference,* (pp. 3021-3031). Orlando, FL

Exemplary Characteristics:

- Training effort included multiple methods; particularly in several platforms that were related to those of interest (i.e., videogame-based training; immersive scenario-based simulation).
- Assessments were coordinated across the battery of training platforms/interventions and included both platform-specific and cross-platform assessment strategies.

Key Takeaway:

The training effort provided a demonstration of how pairing learning domains (e.g., knowledge, comprehension, individual skill, and collaborative team-based skills) to the most appropriate learning technology and introducing domains in an iterative fashion can be effective for enhancing learning outcomes. In particular, the results of the evaluation showed significant learning outcomes when comparing the multi-platform training to an enhanced version of the standard training alone. In addition, the integration of multiple methods of learner assessments (both individual and collective), across multiple technology-mediated training interventions creates a comprehensive picture of learner abilities associated with the full range of underlying competencies that drive task performance in multiple contexts (i.e., individual and team performance). The effort also demonstrated the criticality of a well-developed evaluation strategy to measure training outcomes, and demonstrated the flexibility of using a multiple-platform training strategy to differentially remediate deficiencies within a single training effort.

EXEMPLAR 2: Case-Based Next Generation Cognitive Training Solutions

Citation:

Ross, W. A., & Kobus, D. A. (2011). Case-based next generation cognitive training solutions. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference,* (pp. 2858-2864). Orlando, FL

Exemplary Characteristics:

- Use of SJTs grounded in cognitive task analysis (CTA) to assess underlying cognitive dynamics involved in team-oriented scenario-based virtual world training, partnered with a more qualitative approach (AAR) for feedback and to reinforce learning.

- Training distributed through a 'blended approach' that integrated live instructional training with a fully immersive team-based simulation trainer.
- Results of assessments (both SJT and AAR feedback) designed to 'take home' with units for follow-on programming of training and development activities.

Key Takeaway:

The use of more sophisticated assessment methods, such as situational judgment tests, are particularly effective in assessing higher-level cognitive skills (e.g., decision-making). This is particularly evident when SJT assessments are based on a systematic method of identifying the foundational cognitive requirements of tasks, such as the use of cognitive task analysis (CTA). In this way, the activity linked cognitive training directly to tasks in an operational environment. In addition, pairing quantitative interpretations of results and a more subjective and qualitative assessment of learner outcomes (e.g., AAR) provides broader and more 'user friendly' feedback that allows the learner to use results for targeted self-development activities after training has concluded.

EXEMPLAR 3: Ubiquitous learning website: Scaffold learners by mobile devices

Citation:

Chen, G. D., Chang, C. K., & Wang, C. Y. (2008). Ubiquitous learning website: Scaffold learners by mobile devices with information-aware techniques. *Computers & Education, 50*, 77-90.

Exemplary Characteristics:

- The case included components of mobile, virtual classroom (e.g., interaction with mentors via chat and messaging functions) and live classroom instruction. The system demonstrated how learning outcomes were enhanced through a feedback system, ubiquitously through mobile devices (i.e., feature phone or PDA).
- System incorporated not only assessment of learning (through multiple-choice items) but also tracked learner behaviors (e.g., time spent on training unit, student's learning goals, self-evaluation of content knowledge, and planned timeline to complete course objectives). These were all incorporated into a model of each individual's learning.
- The learning website provided customized (adaptive) learning content based on each student's learning model (i.e., performance on lesson quizzes; time spent on lessons; completion of lessons based on timeline).

Key Takeaway:

The learning website was developed to provide adaptive learning materials based on the student's learning model within a ubiquitous learning environment. The website customizes content to be delivered to any device (PC, laptop, PDA, and cell phones) based on each student's learning model. The student model includes a student's learning preferences and their learning status for every concept in a course. The information-aware system adopts the student model to determine what recommendation (i.e., learning content) should be made and transmitted to a student's cell phone. Additionally, the information-aware system can remind students about scheduled tasks and recommend mentors depending on his/her schedule and achievement of learning concepts. This learning website enhanced students' academic performance, task accomplishment rates, and achievement of learning goals.

Exemplary Element 1: Adaptive Testing

Citation:

Guzman, E., Conejo, R., & Perez-de-la-Cruz, J. (2007). Improving student performance using self-assessment tests. *Intelligent Educational Systems, 22*, 46-52.

Key Takeaway:

Authors demonstrated that students who engaged in the adaptive self-assessments prior to the end of semester exam performed better than students who did not engage in the self-assessment practice exams. They also noted that the frequency of self-assessments demonstrated a weak (but statistically non-significant) linear relationship with final exam scores.

Exemplary Element 2: Continuous Evaluation - Virtual Classroom

Citation:

Coll, C., Rochera, M. J., Mayordomo, R. M., & Naranjo, M. (2007) Continuous assessment and support for learning: An experience in educational innovation with ICT support in higher education. *Electronic Journal of Research in Educational Psychology, 5*, 783-804.

Key Takeaway:

Authors indicated the virtual classroom provided an opportunity to continuously evaluate student learning and served as a platform to supplement in-class learning. Results (although anecdotal) indicated a larger percentage of students pass the course than prior years, with a higher number of students earning an 'A' or 'B' than in past years.

Exemplary Element 3: Extending Testing - Retention Test After 8 Weeks

Citation:

Tucker, J. S., Goodwin, G. A. (2010). Soldier performance following distributed and traditional digital skills training. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conferenc*e, (pp. 2958-2968). Orlando, FL

Key Takeaway:

Skill decay assessed 8 weeks post training was found to be similar regardless of the mode (dL vs. F2F) of training. Authors suggest system cuing during training to aid knowledge accusation and retention.

Exemplary Element 4: Immediately Monitor Learner's States via Recorded Learning Profiles

Citation:

Chen, C.-M., & Hsu, S.-H. (2008). Personalized intelligent mobile learning system for supporting effective English learning. *Educational Technology & Society, 11*(3), 153-180.

Key Takeaway:

The PIMS system continually evaluates each student learner's English vocabulary to ensure appropriately challenging news articles are provided to encourage learning. The system tracks learner's vocab performance and recommends appropriate articles based on the difficulty of the article and whether article has been reviewed by student. The system continually monitors student engagement and will send email reminders to encourage participation of students do not engage the PIMS.

Exemplary Element 5: Competency-Based Assessment of Teams in Collaborative Simulations

Citation:

Rosen, M. A., Salas, E., Silvestri, S., Wu, T. S., & Lazzara, E. H. (2008). A Measurement Tool for Simulation-Based Training in Emergency Medicine: The Simulation Module for Assessment of Resident Targeted Event Responses (SMARTER) Approach. *Society for Simulation in Healthcare, 3* (3), 170-179.

Key Takeaway:

The authors suggest the use of diagnostic measurement tools such as checklists, which can capture 'percent complete/achieved' within each simulated scenario. Alternatively, rating scales could be used, but introduce a more subjective assessment of performance.

Exemplary Element 6: Mentoring and Social Interactions in Virtual Context

Citation:

de Freitas, S., Rebolledo-Mendez, G., Liarokapis, F., Magoulas, G., & Poulovassilis, A. (2010). Learning as immersive experiences: Using the four-dimensional framework for designing and evaluating immersive learning experiences in a virtual world. *British Journal of Educational Technology, 41*(1), 69-85.

Key Takeaway:

Authors utilized Second Life as a platform to provide supplemental mentoring and tutoring for life-long learners. In general, student reactions to mentoring in Second Life were favorable. Authors failed to assess the impact of Second Life above and beyond traditional, F2F mentoring and tutoring or whether Second Life influenced student learning outcomes.

Exemplary Element 7: Narrative Storytelling

Citation:

Hays, M. J., Silva, T. M., & Richmond, T. (2011). Assessing learning from a mixed-media, mobile counter IED trainer. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference*, (pp. 2281-2289). Orlando, FL

Additional Training Content Information:

Content included information on how to react to IEDs. It also provided perspectives from both insurgents and Army Soldiers who have experienced IED attacks in a narrative storytelling fashion. The idea was that, by examining the perspectives, it would demystify IEDs and assist Soldiers in going into 'shock' when they experience an IED attack (and use their counter-IED training). In addition to narrative storytelling (supplied through multi-media), they also examined a number of physical exhibits in the ExCITE modules. For example, one was the interior of a middle eastern home where bomb making had taken place). Finally, there was a virtual simulation game where players navigated an Army convoy and experienced an IED attack. They were supposed to use their learning to respond in the game. Fire teams traded places at being the Army convoy and acting as an insurgent team.

Key Takeaway:

The training program utilized a narrative storytelling approach to convey information via a multi-media video. The idea was that examining different perspectives (via storytelling) would demystify IEDS and allow Soldiers to focus on their training. Results of the training, which incorporated a storytelling video component, demonstrated statistically significant improvement in pre- to post-training scores.

Exemplary Element 8: Gaming to Improve Motivation to Train

Citation:

Mautone, T., Spiker, A., Karp, M. R., & Conkey, C. (2010). Using games to accelerate aircrew cognitive training. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference*, (pp. 1898-1909). Orlando, FL

Key Takeaway:

Results indicated that those who received the GBT version of the training performed significantly better ($p < .05$) on the criterion transfer assessment than those who had received the conventional training, making fewer wording, sequence, and timing errors – thus providing further support for efficacy of GBT.

Exemplary Element 9: Importance of SMEs for Development

Citation:

Okuda, H., Arcaro, L., & Gaught, B. (2011). Understanding the healthcare simulation development lifecycle [Presentation slides]. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference,* (pp. 483-519). Orlando, FL

Additional Information:

Coder's Note: This presentation outlines the specific steps developed as part of the VHA's SimLEARN Healthcare Simulation Lifecycle Model for simulation development. In general, the model contains the following steps:

Phase I – Select a Group of Subject Matter Experts

Phase II – Define Program Requirements
• The overall training objectives, target audience, and clinical scenario

Phase III – Research
• Research results (target audience, organization goals, accreditation requirements, and clinical research)

Phase IV – Define Clinical Content
• Clinical flow diagram of scenario and curriculum spreadsheet

Phase V – Develop Evaluation Plan
• Program evaluation plan

Phase VI – Develop Simulation Template
• Simulation template

Phase VII – Develop Supporting Curriculum
• Supporting curriculum

Phase VIII – Development and Testing of Simulation
• Simulation and individual assessment instruments

Phase IX – Implement Simulation, Train, & Evaluate
• Individual assessment and overall program evaluation data

Phase X – Prepare Evaluation Report for Stakeholders
• Program evaluation report
Particular emphasis is made, for development of the training, on the importance of utilizing a strong group of SMEs. It is recommended that they include technical SMEs (in this case a variety of clinicians), instructional designers, simulation developers, and those knowledgeable in program evaluation. The recommendation is that the panel be involved at all crucial points of the development process.

Particular emphasis given to defining 'the clinical content.' This involves mapping learning objectives to a scenario through development of a flow diagram (which also identifies the crucial learning path...the sequential series of actions that leads to the correct execution of the learning objectives in the presented scenario).

Key Takeaway:

SMEs should be incorporated into the development of training programs. SMEs should be leveraged to identify specific learning outcomes and to develop instructional designs that meet learning outcomes.

Exemplary Element 10: Game for Assessment Function

Citation:

McDowell, P., Johnson, R. E., Freeman, J., Roberts, M., & Horn, Z. (2011). Building a game to educate senior officers in counter-piracy. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference*, (pp. 2301-2311). Orlando, FL

Additional Assessment Information:

The assessment system described is principally focused on an intelligent tutoring-type system. This system utilized for informing the intelligent tutoring capability and estimating performance (to provide feedback to instructors and students) is a proprietary system developed by Aptima. This being the case, very little information was provided in the article about the specifics about how performance was assessed or decisions were made by the intelligent tutoring system for programming scenarios to present to the student. That said, there was discussion on how performance parameters can be programmed into each scenario. Instructors (or IT experts working for instructors) can set scenario parameters that measure actions taken by the student in the scenario or physical changes made (e.g., directions for movement of ships and helicopters) to score performance. Rules are developed to associate to these parameters.

Key Takeaway:

Authors describe software automatically calculates performance measures and provide assessments of decision making during training. These assessments are then published to a database where they are available for inclusion in a variety of reports and can be provided to the trainee as feedback during debrief.

Exemplary Element 11: Rapid Assessment Techniques

Citation:

Kalyuga, S. (2006). Rapid assessment of learners' proficiency: A cognitive load approach. *Educational Psychology, 26*(6), 735-749. doi: 10.1080/01443410500342674

Key Takeaway:

Although this study used a K-12 sample, it provides an exemplary element related to rapid assessment. Rapid assessment using computerized software is able to pinpoint learner weaknesses in proficiency and reduce guessing behaviors. Learner performance demonstrated a significant correlation ($r = .66$; $p < .01$) between the traditional paper and pencil tests score and rapid assessment test score.

Exemplary Element 12: Multiple Testing During Training

Citation:

Lancaster, J. W. & McQueeney, M. L. (2011). From the podium to the pc: A study on various modalities of lecture delivery within an undergraduate basic pharmacology course. *Research in Science & Technological Education, 29*(2), 227-237.

Key Takeaway:

Although it is not uncommon to see training programs possess only a single post-test assessment of knowledge and/or abilities, the method of employing a single post-test assessment lacks the rigor needed to understand the learning process and track trainees across a training platform. This study used multiple quizzes, participation assignments, and a final examination to compare students in a traditional, online, and blended course.

Exemplary Element 13: Computerized Materials Offered with Relevant Instructional Supervision

Citation:

Pereira, J. A., Pleguezuelos, E., Merí, A., Molina-Ros, A., Molina-Tomás, M. C., & Masdeu, C. (2007). Effectiveness of using blended learning strategies for teaching and learning human anatomy. *Medical Education, 41*(2), 189-195.

Key Takeaway:

This study developed and used a virtual campus created with Macromedia Dreamweaver to allow students to have access to relevant computerized learning materials. Additionally, students had the ability to receive assistance from instructors and others classmates via a virtual forum. This relates to the ALC 2015 characteristics of a Learner-Centric 2015 Learning Environment that recommends access to digital learning resources.

Exemplary Element 14: Application of an Existing Model of Teaching and Learning to a Virtual World Environment

Citation:

Salmon, G. Ming, N., & Palitha, E. (2010). Developing a five-stage model of learning in second life. *Educational Research, 52*(2), 169-182.

Key Takeaway:

Although the use of Virtual Worlds in training and development is increasing, there still exists paucity in the research and literature regarding best practices for the development and implementation of a virtual worlds platform. This study describes a method that allows trainees to progress through stages of platform familiarity and exposure to ensure that learning can occur in a virtual world.

Exemplary Element 15: Computer-Based Higher Order Cognitive Skills Training

Citation:

Bolstad, C. A., Endsley, M. R., Costello, A. M., & Howell, C. D. (2010). Evaluation of computer-based situation awareness training for general aviation pilots. *The International Journal of Aviation Psychology, 20*(3), 269-294

Key Takeaway:

This study provides an overview of training aimed at developing higher level cognitive skills via computer-based modules to improve situational awareness and attention sharing for pilots. The extent that this training was effective was determined by virtual flight simulator performance. Thus, the study provides an example of identifying, developing, and assessing higher level cognitive skills.

Exemplary Element 16: Formative Assessments In Mobile Learning

Citation:

Hwang, G-J., & Chang, H-F. (2011). A formative assessment-based mobile learning approach to improve the learning attitudes and achievements of students. *Computers & Education, 56*, 1023-1031.

Key Takeaway:

This study assesses the process of learning within a mobile learning platform. Although this study used 5th-grade students, it was an innovative use of mobile technology, which manipulated the way learners interacted with mobile devices. The value of this research is the clever integration of mobile learning into the context of a physical environment, which can apply to learning about terrain, selection of position, and other outdoor training activities.

Exemplary Element 17: Assessing Simulation-Based Team Training

Citation:

Rosen, M. A., Salas, E., Weaver, S. J., Lazzara, E. H., King, H. B., & Robinson, D. (2010). [Presentation slides]. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference,* (pp. 974-1060). Orlando, FL

Key Takeaway:

This presentation discusses a useful framework (Event-Based Approach to Training or EBAT) for the development of team based assessments across all training platform. EBAT links assessments to targeted behaviors within team scenarios based on prescribed training objectives. The authors also discuss imperative decision points in the assessment design process. This framework and the related design decision points are valuable in that they provide a guide for creating valid and reliable team assessments in a collaborative simulated environment. Such an environment is a focal point of interest, as outlined in the Performance Work Statement (PWS).

Exemplary Element 18: Transfer Practical Exercise To Game Environment

Citation:

Topolski, R., Leibrecht, B., Cooley, S. Rossi, N., Lampton, D., & Knerr, B. (2010). *Impact of game-based training on classroom learning outcomes*. (Technical Report 2010-01). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences (DTIC No. ADA531677).

Key Takeaway:

Soldiers received training via Army Advanced Leaders Courses (ALCs) and then participated in practical exercises (PEs) in either a videogame based environment or terrain boards. The group receiving the PE training via videogame-based training outperformed the non-videogame group. Thus, this element of the study is exemplary in that it provides guidelines for the development and assessment of performance within a videogame-based task in a military context.

Exemplary Element 19: Adaptive Learning Model

Citation:

Kalyuga, S. (2006). Assessment of learners' organized knowledge structures in adaptive learning environments. *Applied Cognitive Psychology, 20*(3), 333-342. doi: 10.1002/acp.1249

Key Takeaway:

Although the study includes an 11th grade population, it demonstrates exemplary use of adaptive training via adaptive assessment techniques and pre-test proficiency to determine subsequent computer-based training content. As a result, knowledge gains and instruction time demonstrated improvements compared to a non-adaptive group. Thus, the value of this study is demonstrated in its ability to highlight one method of using adaptive assessment in a computer-based environment.

Exemplary Element 20: Assessing Communication In Online Settings

Citation:

So, H. J. (2009). When groups decide to use asynchronous online discussions: Collaborative learning and social presence under a voluntary participation structure. *Journal of Computer Assisted Learning, 25*(2), 143-160. doi: 10.1111/j.1365-2729.2008.00293.x

Key Takeaway:

This study addresses issues relevant to distributed learning and in particular, tools available to facilitate learner interaction within a virtual classroom setting. The research addresses use of on-line discussion forums in a self-directed or voluntary context. The study reveals that the degree

of adoption of on-line discussion forums in a collaborative learning task are impacted by initial experiences and the perception of the utility of the on-line tool used. The findings also suggest that the influence of group decision-making on the issue of whether to use the tool impacted the level of individual participation.

Exemplary Element 21: Assessing Chat Room Content

Citation:

Wang, A. Y., Newlin, M. H., & Tucker, T. L. (2001). A discourse analysis of online classroom chats: Predictors of cyber-student performance. *Teaching of Psychology, 28*(3), 222-226.

Key Takeaway:

This study provides another method for assessing chat room content. Thus this study demonstrated the exemplary element of assessing chat room content. The assessment method used was Discourse Analysis (DA). DA encompasses the use of chat log data as well as rater coding to assess chat room based content.

Exemplary Element 22: Setting Learning Objectives In Virtual Reality

Citation:

Tichon, J. (2007). Training cognitive skills in virtual reality: Measuring performance. *CyberPsychology & Behavior, 10*(2), 286-289.

Key Takeaway:

This article yields an exemplary element in that it provides guidance in developing virtual reality training and assessments. Following these guidelines will help ensure that developers do not make the mistake of focusing on the novelty of emerging platforms and, instead, focus on linking assessments and content to training objectives.

Exemplary Element 23: Assessing Performance In Virtual Environments

Citation:

Bowling, S. R., Khasawneh, M. T., Kaewkuekool, S., Jiang, X., & Gramopadhye, A. K. (2008). Evaluating the effects of virtual training in an aircraft maintenance task. International *Journal of Aviation Psychology, 18(*1), 104-116.

Key Takeaway:

This study provides a method for assessing performance in a virtual environment. Here, objective measures, such as number of errors found and search time, were used to assess performance in the virtual search task. Although these measures seem parsimonious, it highlights the potential to assess performance using global, objective measures captured electronically.